

AntConc

テキストファイルの読み込み

File > Open File(s) as Quick Corpus

基本メニュー

KWIC

- ・キーワード検索
- ・前後の文脈のどこで並べ替えるか

Plot

- ・キーワードがデー全体のどこで使われているか図示
- ・Dispersion 分散率

File View

- ・ファイルを開いて該当語を一つずつ見ていく

Cluster

- ・連語表現の抽出

N-Gram

- ・連語表現リスト作成

Collocate

- ・二語の結びつき関係

Word

- ・語彙リスト作成

Keyword

- ・Reference corpus と比べて、特徴語のリスト表示

Wordcloud

- ・高頻度語のグラフィック表示
- ・Word メニューでリストを作っておく
- ・stopword list を指定しないと機能語ばかり出てくる（下記参照）

Tips

Stop words

- ・ 高頻度の機能語などのリストのこと
 - ・ 内容語を見たいときには、こうした高頻度語を除いて分析したいので
- ・ NLTK の stopword list などを取ってきておく

NLTK's list of english stopwords

<https://gist.github.com/sebleier/554280>

<https://gist.github.com/sebleier/554280/raw/7e0e4a1ce04c2bb7bd41089c9821dbcf6d0c786c/NLTK's%20list%20of%20english%20stopwords>

- ・ Settings > Global Settings > Tool Filters > Hide words in file
 - ・ Add File で、取ってきて置いたファイルを指定 > Apply

Corpus Manager で、コーパスデータの管理

- ・ File > Open Corpus Manager
- ・ ファイルを読み込むと、temp.db ができる。これを保存しておかないと、後で使えない。
- ・ 右側で、表示しておき、Save (to file) で、db ファイルで保存
- ・ それ以降は、Add Database File(s) で読み込む

Target Corpus vs. Reference Corpus

- ・ 二種類のコーパスの比較
- ・ Corpus Manager で管理
- ・ どちらを Target、どちらを Reference と指定する。
 - ・ 左ボックスの下の「Open as Target/Reference Corpus」で設定
- ・ 両方読み込んだうえで、
 - ・ File メニューで、Swap Target/Reference Corpora で切り替える