

Collocation

Introduction to Methods in Corpus Linguistics

Day 4

Kristopher Kyle

Yonsei University

Questions?

Dealing with XML texts

- XML texts can also be used in AntConc (depending on the particular XML structure)
- Lets open BNC Baby (which is comprised of XML files) and run some analyses
 - Global Settings -> Tags -> Hide Tags
 - Global Settings -> Files -> XML or ALL

Corpus analyses

In a basic sense, most corpus analyses are used to help make sense of large amounts of language data by organizing and/or filtering it

What do the following analyses highlight?:

- Frequency lists
- Keyword lists
- Clusters/bundles/*n*-grams

Corpus analyses

- In what way(s) can we subsequently analyze/organize our corpus findings?
 - semantically
 - syntactically/grammatically
 - functionally
- Which is the best type of corpus analysis?

Collocation

- From a conceptual point of view, what is a collocation?
- What are some parameters that must be considered when measuring collocations?
 - Left and right span
 - Statistical procedure
 - MI
 - T
 - Other variants
 - Minimum frequency threshold
 - Minimum statistic threshold

Extracting collocations with AntConc

- Lets look at the collocations of *report* in the news section of BNC Baby
 - Span: 4L, 4R
 - Statistical procedure: MI (Tool Preferences -> Collocates -> Collocate Measure -> MI)
 - Minimum frequency: 10 per million
 - Minimum statistic threshold: 3
- Choose a collocate that is a content word which occurs on both the right and the left of the word at least 5 times
- Determine the nature of the collocation
 - semantically
 - syntactically/grammatically
 - functionally

Extracting collocations with AntConc

- Now, lets use another commonly used statistic to investigate collocation
- Lets look at the collocations of *report* in the news section of BNC Baby
 - Span: 4L, 4R
 - Statistical procedure: T
 - Minimum frequency: 10 per million
 - Minimum statistic threshold: N/A

Time to work on Mini Project 2

- You may now take some time to work on your collocation analyses for Mini Project 2
- Don't forget to download the Anaconda version of Python 3!!!