Keyword Analysis

Introduction to Methods in Corpus Linguistics

Day 2

Kristopher Kyle

Yonsei University

Primary research designs in CL

The following are broad generalizations (and overlap to some degree):

Descriptive

- How does X linguistic feature function in a language?
  - Also, what is the distribution/patterning of seemingly analogous forms? (e.g., dative alternation in English)
  - What are the most commonly used realizations of X linguistic feature in a language/particular context (e.g., intensifiers in academic discourse)

Comparative

- How does the use of X linguistic feature differ based on Y individual/group feature?
  - differences in language use across modes, genres, registers, sociolinguistic communities, gender, sexuality, etc.
  - differences in language use across proficiency levels

Trajectory oriented

- How do learner's use of X linguistic feature change over time (i.e., during development)?
- How has the use of X feature changed over time (i.e., over the past 400 years)?

CL-focused journals:

- Corpus Linguistics and Linguistic Theory
- International Journal of Corpus Linguistics
- Corpora
- International Journal of Learner Corpus Research
- (Corpus studies are also published in a wide variety of journals related to both general and specific issues in linguistics and applied linguistics)

Activity 1

- **Research Question:** How is the word "bank" used in American English?

- **More precise research question**: What is the semantic and grammatical distribution of the word "bank" in the Brown corpus

- **Method:**
  - Get a sample of 100 occurrences of the word "bank" in the Brown corpus
  - Determine the uses/meanings of "bank", including
    - part of speech (e.g., verb, noun, adjective, etc.)
    - meaning

Activity 2

1. What are the most frequent words in the Brown corpus?
2. What are the most frequent content words in the Brown corpus?
3. What are the most frequent content lemmas in the Brown corpus?

Keyness analysis

- What does a keyness analysis tell us?

- A keyness analysis highlights words that occur unusually frequently (or unfrequently) in a particular corpus as compared to some reference corpus

- This information can be used for a variety of purposes, but can tell us what a corpus is "about"

- More precisely, it can tell us which words tend to be common in a particular corpus compared to "normal" usage

Conducting keyness analysis in AntConc

- We will conduct a keyness analysis using the Brown corpus.
- The "government" section will be our target corpus, and the other 14 sections will be our reference corpus

Corpus design brainstorming

1. Target language use domain
   a) Reference corpus of …
   b) Specialized corpus of …
2. What information do you want to extract from a corpus?