

## **Assignment Guide for Mini-Project 1**

### Introduction to Methods in Corpus Linguistics

#### **Purpose of project**

To demonstrate functional knowledge of common analyses in corpus linguistics including:

- a) Compiling a specialized corpus,
- b) generating normed word frequency lists,
- c) generating a keyword list for your specialized corpus

#### **Procedure:**

##### ***Step 1: Compiling a specialized corpus***

You should collect a (relatively small) corpus that is representative of a particular language use domain. You can use pre-existing corpora as part of your corpus (if an appropriate one exists), but you also need to collect some new data as well.

Be sure to report:

- a) The language use domain that your corpus represents
- b) The number of words in your corpus
- c) The number of documents included in your corpus
- d) How you collected the data
- e) Benefits and limitations of your corpus

##### ***Step 2: Generate normed word frequency lists***

Using AntConc, generate a lemmatized word frequency list. In separate tables, report the 20 most frequent words (all words) and the 20 most frequent content words (you will need to use a stoplist for this). Be sure to include:

- a) Their rank,
- b) Their raw frequency
- c) Their normed frequency

##### ***Step 3: Generate a key word list***

Using AntConc, generate a key word list using the log-likelihood method (4 term). In a table, report the 20 most frequent words (all words). For each word, be sure to report:

- a) Rank,
- b) Normed frequency
- c) Keyness value
- d) Effect size

***Step 4: Investigate use of keywords in each corpus***

Often, word use varies by language use domain. Using AntConc, choose two of your keywords (make sure that the words you choose are content words) to investigate. Explore the use of each of these words using the concordancing tool in AntConc. Report the use of each word in your specialized corpus, being sure to include examples in your description. For words that are used in multiple ways, you only need to describe the 2-3 most common uses. Compare the use of each word with the way(s) that it is used in the reference corpus.

## Mini-Project 1

### Introduction

<insert introduction here – provide an overview of the project and indicate why the analysis you are conducting might be interesting/useful>

### Target Corpus

<Describe characteristics of the target corpus here. This is the one that you designed and collected. Be sure to outline the target language use domain and the purpose of the corpus.>

### Reference Corpus

<Describe characteristics of the reference corpus here. Most likely this will be the Brown corpus, but you may optionally choose a different corpus.>

### Frequency List for Target Corpus

<Include a table here with the 20 most frequent words and their rank, raw frequency, and normed frequency (per million words). Additionally, submit a supplementary file with the complete frequency list when you submit your report.>

### Keyness Analysis Method

<Describe how you conducted the analysis here, including the settings used>

### Keyness Analysis Results

<Report your keyword list here>

### Word Use Analysis

Introduce the two words that you will further investigate.

**Uses of *word one* in target corpus.** <Insert discussion of uses (with examples from the corpus) here>

**Uses of *word one* in reference corpus.** <Insert discussion of uses, being sure to compare the uses in your target corpus with the uses in the reference corpus>

**Uses of *word two* in target corpus.** <Insert discussion of uses (with examples from the corpus) here>

**Uses of *word two* in reference corpus.**

### Conclusion

<summarize your results here>