

Assignment Guide for Mini-Project 4
Introduction to Methods in Corpus Linguistics
(updated 2021-07-22)

Purpose of project

The purpose of this project is to take what you have learned during the course and apply it to a research question (or small set of questions) of your choosing. You can expand on one of your previous mini-projects if you choose, or you can conduct a new project. Also note that mini-project 4 could also involve the analyses we learned during the last three days of class (e.g., dependency bigrams).

The biggest difference between Mini-Projects 1-3 and Mini-Project 4 is that you will have to more clearly outline the motivation for your project (including some reference to previous work) and the criteria for what constitutes an “appropriate” corpus will be stricter.

Step 1: Project Outline (Due Monday, July 26th before class)

Your project outline should have the following format:

1. Introduction

<Brief description of research gap (why is this research project interesting/important from a practical or theoretical perspective?)>

2. Research Question(s)

<Outline the research question(s) for your project here>

3. Method

Linguistic Analysis

<outline the linguistic features you will be investigating (e.g., words, lemmas, pos_lemmas, ngrams, dependency bigrams, etc.)>

Corpus Analysis

<Outline corpus analysis techniques here (e.g., keyness, collocation, etc.)>

Corpus/Corpora

<Outline the characteristics of the corpus/corpora you will use for your analyses>

4. Expected results

<Provide a brief description of your expected results and/or the potential implications of your findings>

Step 2: Completed Project (Due Monday, August 2nd, 11:59pm)

The format of Mini-Project 4 will vary by person/group. However, it should have the components below (note that the length of Mini-Project 4 should be between 3000 and 4000 words).

1. Introduction

- a) this should be 1-3 paragraphs (approximately 250-350 words)
- b) introduce research area
- c) introduce research gap
- d) briefly introduce how your project addresses the research gap

2. Condensed Literature Review

- a) this should be 3-4 pages (approximately 1200 words)
- b) identify key themes/issues in the literature (these will likely be your headings)
- c) discuss what the field has hypothesized and/or found with regard to these themes/issues

3. Research Question(s)

- a) Include 1-3 clear and concise research questions

4. Method

The Method section should clearly outline what you did. After reading your Method section, I should be able to conduct the same analyses (and get the same results).

Linguistic Analysis

- a) outline the linguistic features you will be investigating (e.g., words, lemmas, pos_lemmas, ngrams, dependency bigrams, etc.)

Corpus Analysis

- a) Outline corpus analysis techniques here (e.g., keyness, collocation, etc.) and the choices made (statistic, etc.)

Corpus/Corpora

- a) Outline the characteristics of the corpus/corpora you will use for your analyses

5. Results

- a) Provide the results of your analyses. This will likely include some tables (e.g., the top hits in your analysis).
- b) Also provide the results of your follow up analyses (e.g., your more fine-grained analysis of the use of the linguistic items you investigated).

6. Discussion and Conclusion

- a) Interpret your results in light of previous research (how did your findings align with or diverge from previous literature) and/or discuss the practical applications for your findings.
- b) Outline the limitations of your study (every research project has limitations that potentially affect the interpretation of the results)
- c) Outline areas for future research (What further analyses would be useful? How could future research address the limitations of your study?)
- d) Summarize the implication of your study/findings