October 3, 2016

**Introduction**

The objective of the assignment is to define a relatively complex term for an audience that is assumed to have no ‘technical knowledge.’ This assignment will use three types of definitions: parenthetical, sentence and expanded. The following paragraphs will sample the three methods for writing an in-text definition, as well as when it is appropriate to use which definition method, depending on the technical expertise of the audience.

**Term:**Corpus analysis

**Parenthetical Definition**

Corpus analysis (a technique for discourse examination) is a comprehensive way to study linguistic phenomena and explore characteristics of language usage.

**Sentence Definition**

Corpus analysis is a text-based methodology used for in-depth study of language and discourse. Corpus analysis helps to reveal linguistic taxonomies, such as value marking (indication of value e.g. good or bad), speech modality (the writer’s degree of confidence in their claims), and lexical bundle usage (grouping of specific words).

**Expanded Definition**

What is corpus analysis?

Corpus analysis is the methodology of studying large databanks of text with computer software in order to draw conclusions about the usage of language. It helps to reveal characteristics of language usages based on “age, gender, level of education, and socioeconomic background; place and time of a communicative event” (Hasko 1).

How is a corpus different from a word database
A corpus is not simply a word database but is a collection of systematically selected text. The text is often restricted to certain type or time span which will be used to study a specific supposition or premise (Nesselhauf 2).

Where did it come from?

Corpus analysis is a modern methodological addition to the field of linguistics. It began around the 1950s with the availability of computers for analysis of large sets of text (McEnery). The very first electronic corpus was The Brown Corpus of Standard American English, which consists of 1 million words from American English texts printed in the 1960s (Corpora Project). In the early history of corpus-based analysis there was opposition because its view of language was seen as too empirical (McEnery). However, since then corpora have proved to be a very useful tool in linguistic theory.

What is corpus analysis for?

The main objective of corpus analysis is to discern patterns that occur in language through analyzing actual language usage (Krieger). It is used to study how meanings and functions interact with language used to communicate (Hasko 1). To better understand what corpus analysis is it may help to breakdown the two main components:

1. First there must be a **hypothesis** or interest in finding a pattern in a specific areas of discourse. This interest will guide the direction of study and clarify the type of text needed.
2. Once a direction for the study, **a corpus** must be created or found. A corpus is a systematic collection and compilation of text, from written or spoken discourse. The text collection is used to create a databank of natural language usage in a specific context.
3. After compiling a corpus, the text is put into a **concordance software** program, such as Antconc. The software allows the user to analyze the corpus by searching the text for instances of words or phrases relevant to the hypothesis. An example of how the software works can be seen in the figure (visual 1) below. This figure demonstrates a search in the concordance software. The highlighted text represents the word or phrase being searched and the number on the right indicates the number of occurrences of that word or phrase.

Visual 1

Image take from page 16 of Nesselhauf’s article.

Example:
An example of how corpus analysis might be used is to compare of the writing of men and women within a discipline. This comparison could be done by collecting academically published articles from women in a specific discipline to create one corpus and then creating a second corpus of academically published articles from men of the same discipline. Comparing the two corpora will highlight the variances between men and women within a specific linguistic area.

In a corpus analysis I conducted myself I looked at the comparison between male and female academic research articles across multiple disciplines. My analysis showed a significant difference between male and female language in terms of self-mention and use of boosters (confidence indicators such as “clearly” or “obviously”). Men tended to show dramatically increased usage in both areas.

The conclusions I made based on my corpus findings were that male usage of self-mention reveals that the male authors of my corpus, more than the females, confidently reflect on their own opinions and a stronger desire to validate their work as distinctly their own. Correspondingly, the variance in usage of boosters also implies that men are more likely than women to display certainty and confidence in their work. Overall, the gender difference in self-mention and boosters demonstrates that across disciplines female academic authors lack confidence in the significance and validity of their claims, when compared to their male counterparts.

Works Cited

Corpora Project. "Introduction: Corpus Linguistics." University of Essex. N.p., 1998. Web. 25 Sept. 2016.

Hasko, Victoria. "Qualitative Corpus Analysis." The Encyclopedia of Applied Linguistics. 2012. Web. 25

Sept. 2016.

Krieger, Daniel. "Corpus Linguistics: What It Is and How It Can Be Applied to Teaching." TESL Journal

Archives. 2003. Web. 26 Sept. 2016.

McEnery, Tony, and Andre W Hardie. "History of Corpus Linguistics." Oxford Handbooks. N.p., 2015.

Web. 24 Sept. 2016.

Nesselhauf, Nadja. "Corpus Linguistics: A Practical Introduction." University of Heidelberg. 2005. Web.

25 Sept. 2016.

Wagner, Hannah. “‘Women Should Be Seen Not Heard’: Analyzing Gender Differences in Academic

Discourse.” ASTU 400. Print.