**Corpus Analysis**

A. Overview of corpora

1. Definition of corpus = A large body of naturally occurring texts

2. Usually involve sampling BUT need not

3. May be general and involve many kinds of texts or specialized sets of particular kinds

4. Many corpora have features of their texts tagged

B General Corpora (examples)

1. British National Corpus -100 million word texts

* 90% written (national and regional newspapers, fiction, leaflets, letters)
* 10% spoken (conversations, meetings, telephone calls)
* texts tagged for parts of speech, genres & other features
* Focuses on British English expression in late 20th century

2. Open American National Corpus — available for free at <http://www.anc.org/>

* 15 million words, subset of 22 million American National Corpus
* American English written & spoken, produced since 1990

C. Specialized Corpora = oral or written texts reflecting discourse in a specific domain

1. Call-Home American English —120 calls originating in North America, 90 to other countries, 30 within.

2. Teubert’s major written Catholic doctrines

D. Key Analytic Terms

1. Frequency & type/token ratio

2. Key word in context—also labeled concordance (usually +/-5 words on either side)

3. Collocation analysis —words that occur most frequently with a term

D. Link to Discourse Analysis

1. Most often used in combination with critical discourse approaches

2. Offers a tool to anchor claims about typical or atypical

**Workshop: Analyze Judge Stavley’s Case**

1. How many different words are used in this trial?
2. How many total words?
3. What is the type-token ratio?
4. What 3 words occur most often? Identify the 2-3 most frequent legal-trial terms used.
5. How often does the word, “appeal” occur? What can you say about its usual context in this trial? How often does “if” appear within 5 words of appeal?
6. How often does “abuse” occur?
7. If you run “violence” as a key word in context, how many instances are there? What is its main substantive collocate?
8. What can you tell me about the word “well”?
9. Describe a question you could ask about this corpus and answer it.
10. Identify an interesting question you could address if you had 3 trials from 3 different judges.