Chapter 1 – Corpus Linguistics

1.1 In this exercise you will get acquainted with the British National Corpus (BNC), one of the corpora used throughout the book, and with some of the basic concepts of corpus linguistics. The search interface, BYU-BNC, is identical with those of the Time Corpus and the Corpus of Contemporary English (COCA).

Before you do the web exercise, read about concordances and frequencies in section 1.4 and about the BNC in section 1.6.2.

(i) Go to http://corpus.byu.edu/bnc/. NB: during the session you may be asked to register as a user. Do this – it is free of charge. Make a note of your password. Read the site Introduction and then click on “five minute guided tour” to learn more about the corpus.

(ii) Click on the word mysterious in the guided tour. It will then occur under “Search string” in the search panel to the left on the screen. Under “Display”, the “Chart” button should be marked. Under “Section”, “Ignore” should be marked. Now press the “Search” button. The results per section of the corpus will be displayed in a bar chart in the right-hand part of the screen. The chart shows that mysterious occurs most frequently in Fiction, with 27.4 tokens (instances) per 1 million words, followed by Miscellaneous, Newspapers, Academic and Spoken.

(iii) Now click on the bar for Fiction in the chart. The first 100 concordance lines with mysterious will be displayed and you will be able to study the contexts in which the word occurs, e.g. what kind of nouns it modifies.

(iv) Go back to the search panel to the left (it may be partially hidden) and mark “List” instead of “Chart”. Click on “Search” again. Now the total number of occurrences is reported in the right-hand panel. If you click on the word mysterious you get the first 100 concordance lines regardless of subcorpus.

(v) Look at this list of five adjectives: big, hot, superfluous, supreme, tragic. Try to guess which is most frequent, second most frequent etc. in the corpus, and then in which subcorpus they occur most frequently. Then search for each of the words in the manner described above and for each of them note down the total frequency and which subcorpus it occurs most frequently in. Don’t forget to look at the concordance lines to get an indication of how the words are used.

(vi) Try searching for some other words that you are curious about.