A BRIEF INTRODUCTION TO THE BASIC USE OF MONOCONC PRO

The aim of this guide is to reproduce an introductory session using *MonoConc Pro*; however, many features of the program are not touched on in this document.

1. STARTING MONOCONC PRO

To start the program, double-click on the file MonoPro. Once the program is open, a simple screen appears, as shown below in Figure 1. This initial screen looks rather bare, containing only a blank window and two menu items: FILE and INFO.


Figure 1: Initial screen

The INFO menu is always present; it provides access to HELP and to some basic information about *MonoConc Pro*, as well as some contact information for Athelstan. The HELP menu is organised according to topic (see Figure 2). Double-clicking on the appropriate heading will open the topic file, allowing perusal of the associated descriptions.

Figure 2: Help contents

Selecting the FILE menu reveals several commands, one of which is the LOAD CORPUS FILE(S) option, but before making a corpus available for searching it may be necessary to provide appropriate info in the LANGUAGE and TAG SETTINGS menu in order for certain commands to work, such as those controlling tag searches and the suppression/display of part-of-speech tags. We will skip this part. The LOAD CORPUS FILE(S) command prompts the user to choose the directory/files to be loaded ready for analysis.

Once a corpus is loaded, some new menu items related to the analysis and display of the text appear on the menu bar. These are FILE, CORPUS TEXT, CONCORDANCE, FREQUENCY, WINDOW and INFO. In addition, looking at the screen in Figure 3 we see information in the lower left corner relating to the number of the files loaded and in the lower right corner a word count for the corpus.
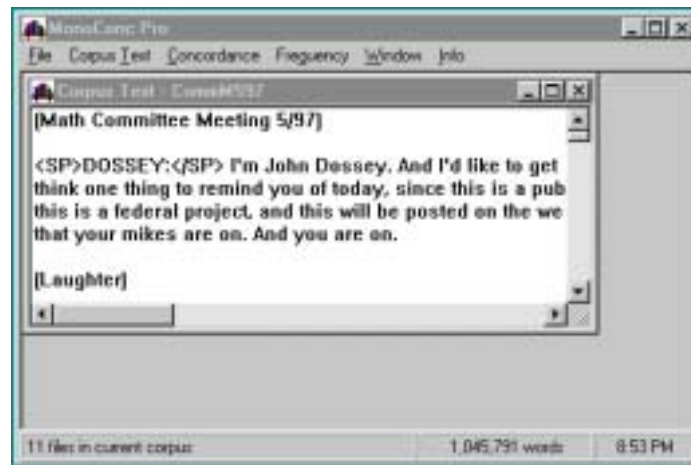
Figure 3: View of a corpus file

2. CORPUS FREQUENCY INFORMATION

Typically, the most frequent content words in a corpus are of interest since they give the best indication of the nature of the corpus. In most cases, all words occurring more than two or three times in the corpus are displayed either in frequency order (from most to least frequent) or in alphabetical order. To find out the distribution of words in the entire corpus, choose CORPUS FREQUENCY DATA and select either FREQUENCY ORDER or ALPHABETICAL ORDER.

Frequency options

Choosing FREQUENCY OPTIONS allows the frequency list to be tailored to fit particular requirements. It is possible to limit the data presented in three main ways, as can be seen from the FREQUENCY AND COLLOCATION OPTIONS dialogue box shown in Figure 4. First, we can set the maximum number of lines in the frequency list (using the MAXIMUM LINES parameter), which means that it is a simple matter to find, for example, the twenty most frequent words or the hundred most frequent words in a corpus. Secondly, a lower frequency boundary can be selected (MINIMUM FREQUENCY). In Figure 4, the minimum frequency has been set to 3, which means that words occurring only once or twice, of which there are many, will be excluded from the list.

Figure 4: Frequency and collocation options

The third main option is to set an upper frequency boundary (MAXIMUM FREQUENCY) that excludes words occurring more often than the set threshold. (The setting of 0, shown above in Figure 4, is used to indicate no upper boundary.)

3. PERFORMING A BASIC CONCORDANCE SEARCH

A concordancer is, at its heart, basically a search program that looks for patterns in the text based on a search query. Simple as this sounds, it can lead to sophisticated analyses of lexical, grammatical and textual structure. The advantages of using a concordance program are that it makes it possible to (i) find rare instances of words or strings; (ii) find strings in the context of other strings, e.g., the instances of *economy* occurring after <title> and before </title>; and (iii) look for particular patterns and then rearrange and concentrate similar instances so that their properties can be revealed.

Let's look at an illustration of what we mean by patterns in the text. Here we can take a mundane but illustrative example and examine the question of what kinds of words typically follow the verb *speak*. Based on our intuitions about English, we might suggest prepositions such as *to* and perhaps *with,* as typically associated with *speak*. To find out what actually occurs, we need to look in a corpus for the pattern [SPEAK + word]. (For the present, we will assume that we are working with an untagged text.)

This formulation is a little misleading, as it is not possible to specify a single term SPEAK which covers the lemma or "word family" *speak*. There are ways of specifying a lemma, but for now let us assume that we have to search for actual words: *speak, speaks, speaking, spoke,* etc. So we will start off with an elementary approach, which is simply to search for all the instances of one verb form, say the base form *speak*.
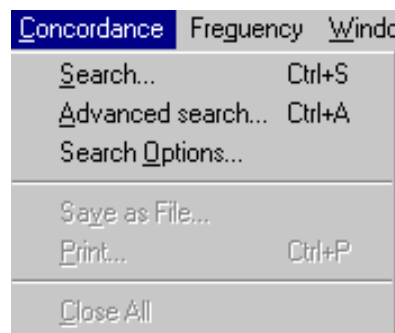

Figure 5: Concordance menu

To do this, we select SEARCH from the CONCORDANCE menu, or enter CTRL-S, (shown in Figure 5) In the text box at the top of the dialogue box that appears (Figure 6) type in the search term **speak** and click on OK (or press enter). The example search queries that can be found below the search box serve to remind us about the format of simple text searches—and the use of wildcards.
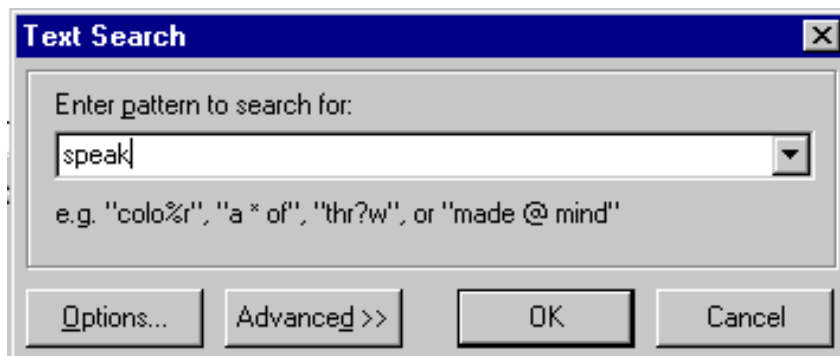

Figure 6: Performing a simple text search

Technical Note: The parameters of the search are determined by the settings in force in GENERAL SEARCH CONTROL in the ADVANCED SEARCH dialogue box and in SEARCH OPTIONS as illustrated in Figure 7. These latter settings cover such things as whether the hyphen is treated a word boundary.

The dialogue box in Figure 7 contains a lot of information. For now, you can perhaps just take away the idea that when you become familiar with the program and want to have more control over the way in which it operates, you will be able to configure it in the way you want. I assume that initially you will perform searches without reference to this information.

Figure 7: Settings in Search options

Let us now look in some detail at what happens during the search process. In our example, the program works through the text looking for the word *speak*. With such a common word, the results should flood in fairly rapidly. The results of the search appear in the concordance results window (Figure 8). In this window, each instance of the search word that the program finds is copied along with a preceding and following context.

Typically, the search word is centred and highlighted so that the instances of the search word line up, as shown in Figure 8. This format is commonly referred to as a KWIC (Key Word In Context) format.
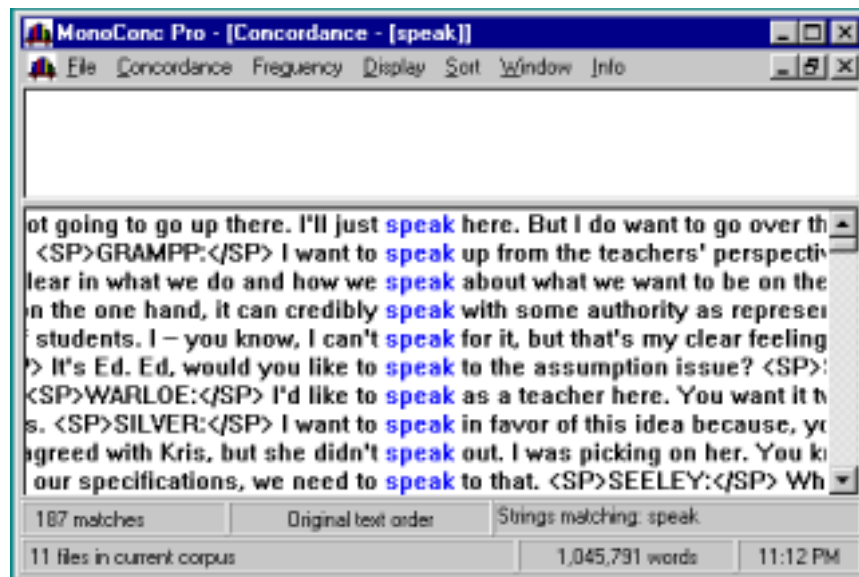

Figure 8: Concordance results in a KWIC format

Once the search is ended, then you can bring to bear the full power of the program to reveal patterns in the results. One way to find out which words are associated with *speak* is to **sort** the instances so that they are in alphabetical order of the word following the search term. The advantage of performing this 'right sort' is that all the instances of *speak to* will be next to each other in the concordance window, as will all the instances of *speak with,* and so on. The easiest way to achieve this ordering is to select 1ST RIGHT, 1ST LEFT from the SORT menu (Figure 9).
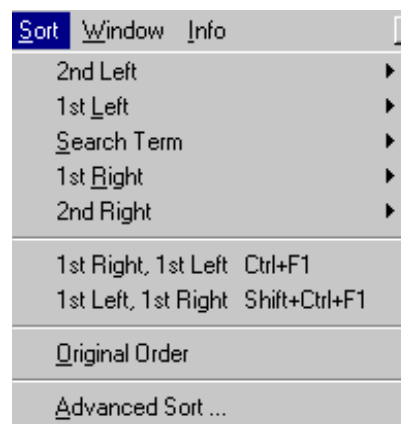

Figure 9: The Sort menu

The program then immediately rearranges the concordance lines to give a more revealing view of the search results, as shown in the sample in Figure 10.
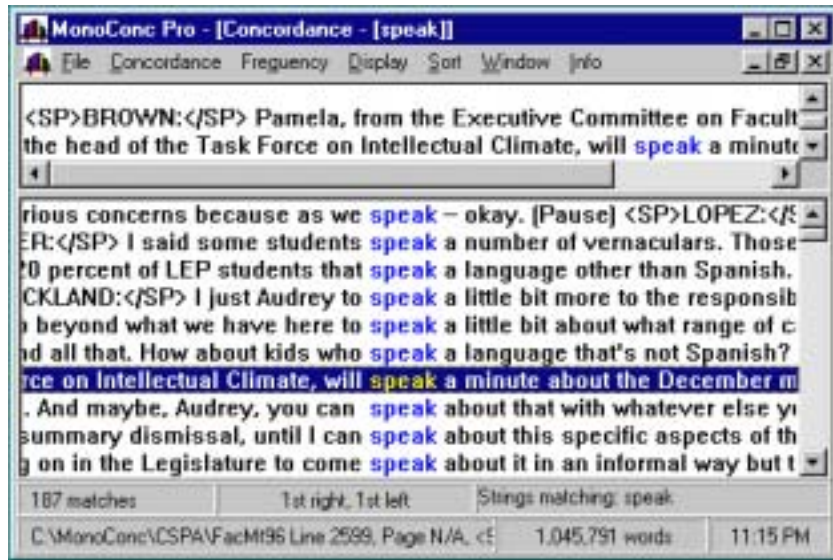
Figure 10: Sorted concordance results

Having ordered the concordance lines in this way, it is very easy to see which words occur with *speak* and with what frequency. Those lines having the same word following *speak* will be clustered together, arranged according to the alphabetical order of the word *following* the search term.

If you scroll fairly quickly through the concordance results, you will discover that the visual patterning created by several identical words surrounding the search word will be striking enough to catch your eye. It is not necessary to focus on the results line by line; you can scan the output quite rapidly.

To see a larger context (i.e., a chunk of preceding and following text) for any concordance line, simply select the line in the results window by clicking on it. This larger context is then displayed in the upper context window. (A rather small context window is shown above in Figure 10.)

Sometimes we need more powerful search options than the simple text search illustrated above. We will not discuss these searches further here, but will simply give a screen shot in Figure 11 of the advanced search dialogue box. The search specified is designed to extract expressions such as *he clawed his way to the top*. The search is for a verb followed by a possessive pronoun followed by the word *way*.
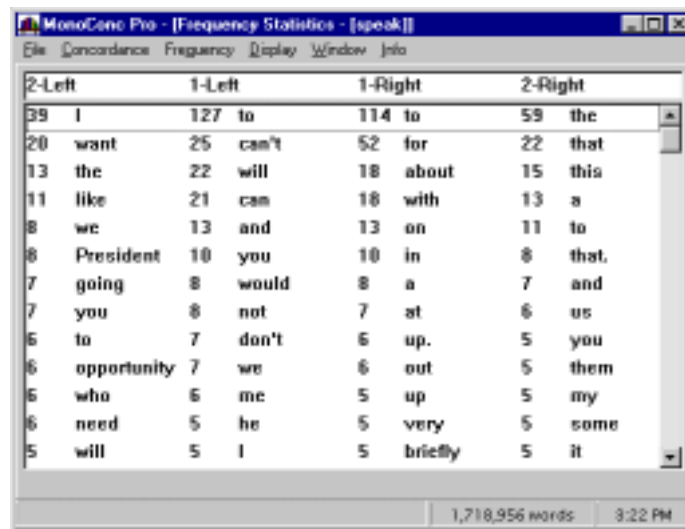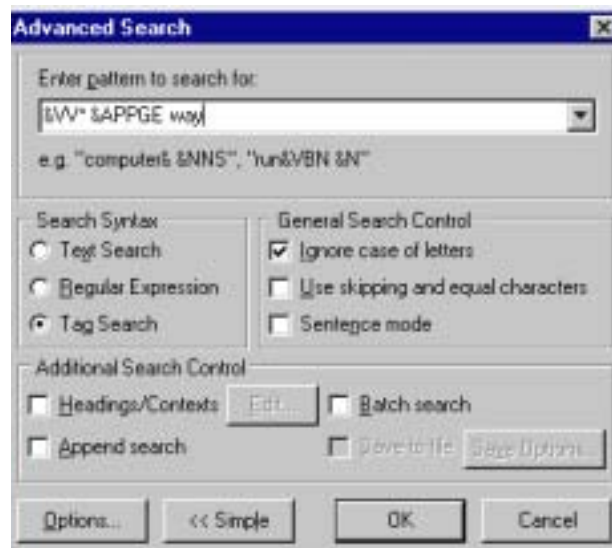

Figure 11: A Tag Search

4. COLLOCATES AND COLLOCATIONS

MonoConc Pro furnishes a variety of frequency statistics, but the two main kinds are corpus frequency and collocation frequency. The command CORPUS FREQUENCY DATA creates a word list for the whole corpus, as described above. Choosing COLLOCATE FREQUENCY DATA from the FREQUENCY menu (or CTRL-F) displays the collocates of the keyword ranked in terms of frequency.

The collocates of a word are its frequent neighbouring words. In *MonoConc Pro,* the collocate frequency calculations are tied to a particular search word and so the frequency menu only appears once a search has been performed. The collocation data produced by the COLLOCATE FREQUENCY DATA command is organised in four columns, with one column for each position surrounding the keyword: 2nd left, 1st left, 1st right and 2nd right. (Thus 1st left refers to the word before the search term and 1st right is the word following the search term.) The columns show the collocates in descending order of frequency.

Above, we examined the simple question of which words frequently follow *speak*. Once we have searched for *speak*, we can select COLLOCATE FREQUENCY DATA and see at a glance what the common words following *speak* are. The collocates of *speak* (in four positions) are given in Figure 12 and we find, not surprisingly, that *to* is the most common word following *speak*, but we also get a sense of the variety of other common collocates of *speak*. The most frequent left collocate of speak is the infinitival *to*. Remember that we searched for the base form *speak* and the grammatical consequences of this choice can be seen in the words showing up in the 1-Left column.

2-Left		1-Left		1-Right		2-Right	
39	I	127	to	114	to	59	the
20	want	25	can't	52	for	22	that
13	the	22	will	18	about	15	this
11	like	21	can	18	with	13	a
8	we	13	and	13	on	11	to
8	President	10	you	10	in	8	that,
7	going	8	would	8	a	7	and
7	you	8	not	7	at	6	us
6	to	7	don't	6	up.	5	you
6	opportunity	7	we	6	out	5	them
6	who	6	me	5	up	5	my
6	need	5	he	5	very	5	some
5	will	5	I	5	briefly	5	it

Figure 12: Collocate frequency table for **speak**

One disadvantage of the simple collocate frequency table is that it is not possible to gauge the frequency of collocations consisting of three or more words. Thus, we cannot tell from the *speak* collocates table how common the phrase *to speak to* is. To calculate the frequency of three word collocations, it is necessary to use the ADVANCED COLLOCATION command in the FREQUENCY menu.

Saving and printing the search results

The output of searches and frequency calculations can either be printed or saved as a text-only file.

5. EXITING THE PROGRAM

To exit *MonoConc Pro*, choose EXIT (CTRL-Q) from the FILE menu.