

## REVIEW OF *MONOCONC PRO* AND *WORDSMITH TOOLS*

<b>Title</b>	<i>MonoConc Pro Version 2.0</i>	<i>WordSmith Tools Version 3.0</i>
<b>Developer</b>		Mike Scot
<b>Platform</b>	PC	PC
<b>Hardware/ System Requirements</b>	Windows 95 or higher	Minimum 80386 processor, VGA display or better, Windows 3.1x or Windows 95, minimum 4 MB RAM (8 MB if used with Windows 95).
<b>Program Information</b>		<a href="http://www.oup.com:8080/elt/global/catalogue/multimedia/wordsmithtools3/">http://www.oup.com:8080/elt/global/catalogue/multimedia/wordsmithtools3/</a>
<b>Publisher</b>	Athelstan <a href="mailto:info@athel.com">info@athel.com</a>	Oxford University Press
<b>Support</b>	On-line help and a small manual	On-line help and an extensive manual
<b>Languages</b>	Can be used with different languages	Can be used with different languages
<b>Audience</b>	Beginning to advanced users	Beginning to advanced users
<b>ISBN</b>	not applicable	0-19-45-92863
<b>Price</b>	US \$85 single user; US \$550 15 user site	51.95 British pounds

### Reviewed by Randi Reppen, Northern Arizona University

The recent interest in corpus linguistics and the use of authentic materials has created a need for software packages that allow teachers and researchers to carry out corpus-based investigations. These corpus-based investigations can be used to augment classroom instruction so that ESL/EFL students are exposed to real language rather than artificial texts and made-up examples. Teachers and researchers can also begin to explore some of the more subtle areas of language use where our intuitions often lead us in the wrong direction.

In this review, I will take a close look at *WordSmith Tools* (Version 3) and *MonoConc Pro* (Version 2), two of the more readily available and reasonably priced packages for working with corpora, in order to contrast the different options that they offer teachers and researchers. As with any software purchase, the needs of the user should play a key role in deciding which program is most appropriate. Both programs include many of the same features, such as the ability to create word lists (in both alphabetical order and frequency order), generate concordance output, and give collocation information. Both programs easily handle large corpora and work with either tagged or untagged texts. As with any software package, the user needs to check the default settings (e.g., minimum or maximum number of hits displayed) to make certain that they are set according to the users' desires. In the following paragraphs, I describe the major features shared by the two programs as well as some of the more specialized features offered by only one or the other.

One of the major innovations of these packages is that they allow users to analyze any collection of ASCII texts. This is in marked contrast to earlier concordancing packages which required the user to build a database of texts before using the program for analyses. This was usually an elaborate process, and sometimes required sending texts to the software author or publisher before the concordancing tools could be used. Further, the database normally could not be modified once it was constructed. Thus, the database needed to be rebuilt any time additional texts were added. *WordSmith* and *MonoConc Pro* differ from these earlier packages in that they allow the user to select any group of texts for analysis every time the system is started. Better yet, additional texts can be added "on the fly," so that the corpus being analyzed can be tailored to directly fit the immediate research questions.

The primary research use of both software packages is to generate concordances, or listings of all the occurrences of any given word in a given text, with words shown in context. Concordance listings can be useful for exploring the use and meanings of specific words. Often when looking at concordance lines, users may want to expand the context so that they can get a better sense of the meaning or use. Here is one area where the two programs differ quite a bit. Both programs allow the user to adjust the settings of the concordance program to display more or less text on the concordance screen. However, *MonoConc Pro* has an additional feature that is especially attractive for researchers: the split screen display allows users to expand the context of an entry line simply by highlighting the line, which displays the fuller context in the upper window (see [Figure 1](#)). In *WordSmith*, the entire display must be expanded or reduced, so the context is expanded for all of the entries being viewed rather than for a single highlighted entry.

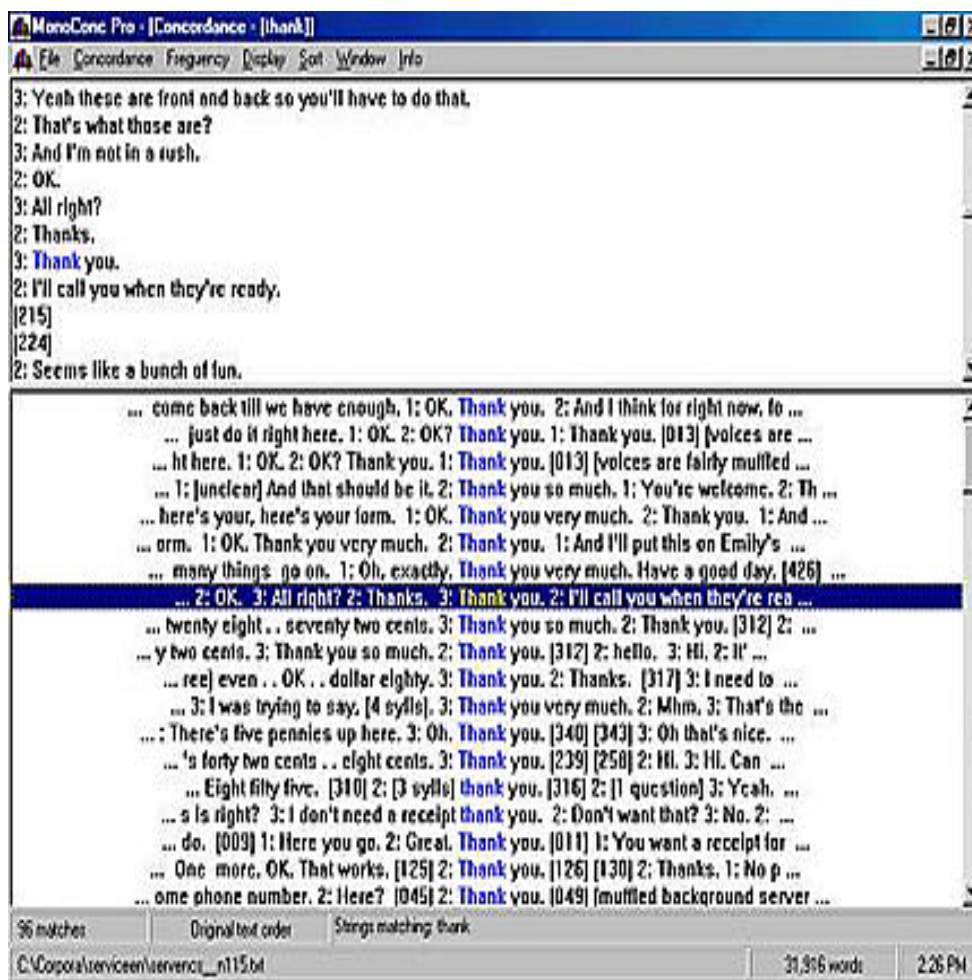


Figure 1. *MonoConc Pro* screen display of concordance lines

Another nice feature of *MonoConc Pro* is that the total number of words in the corpus is always displayed in the lower right hand corner (as shown in [Figure 1](#)). This information is vital for comparisons of texts of unequal lengths, as the normalization of counts of linguistic features, a process that allows such comparisons to be carried out accurately, relies on text length (for more information, see Methodology Box 6 in Biber, Conrad, & Reppen, 1998, pp. 263-264).

Both programs have sort functions that allow users to sort concordance lines in several ways (e.g., by search word, then first word right; or by first word). Sorting words and seeing the collocation

immediately to the left or right of the target word can provide insights on word senses and uses. Another feature found in both programs is the ability to "blank out" target words in the concordance output, which can be useful to teachers for the development of vocabulary activities and cloze tests. By using corpora, rather than teacher-made examples, teaching and testing materials reflect the language found in authentic texts and thus provide learners with more exposure to real language. Concordance displays are quite similar in both programs.

In addition to the functions that these programs have in common, *WordSmith* is able to perform a number of useful tasks that *MonoConc Pro* is not. For example, *WordSmith* can provide information about the distribution of a feature in a single text or across texts. Distributions are shown with a graph that plots the occurrences of the target item in the text or corpus (see [Figure 2](#)). The distribution of a particular lexical or grammatical feature across a text or series of texts can provide interesting information about the text structure and also about how the feature functions across various texts. A similar tool is available in *MonoConc Pro*; however, I was unable to interpret the bar graph display used in *MonoConc Pro*.

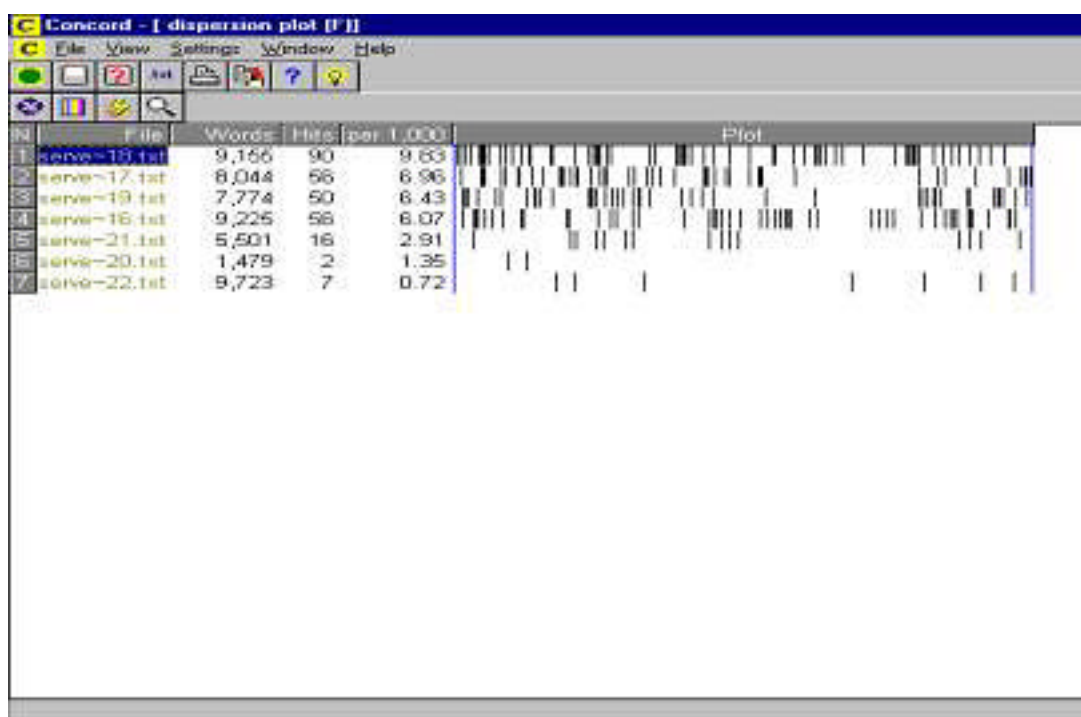
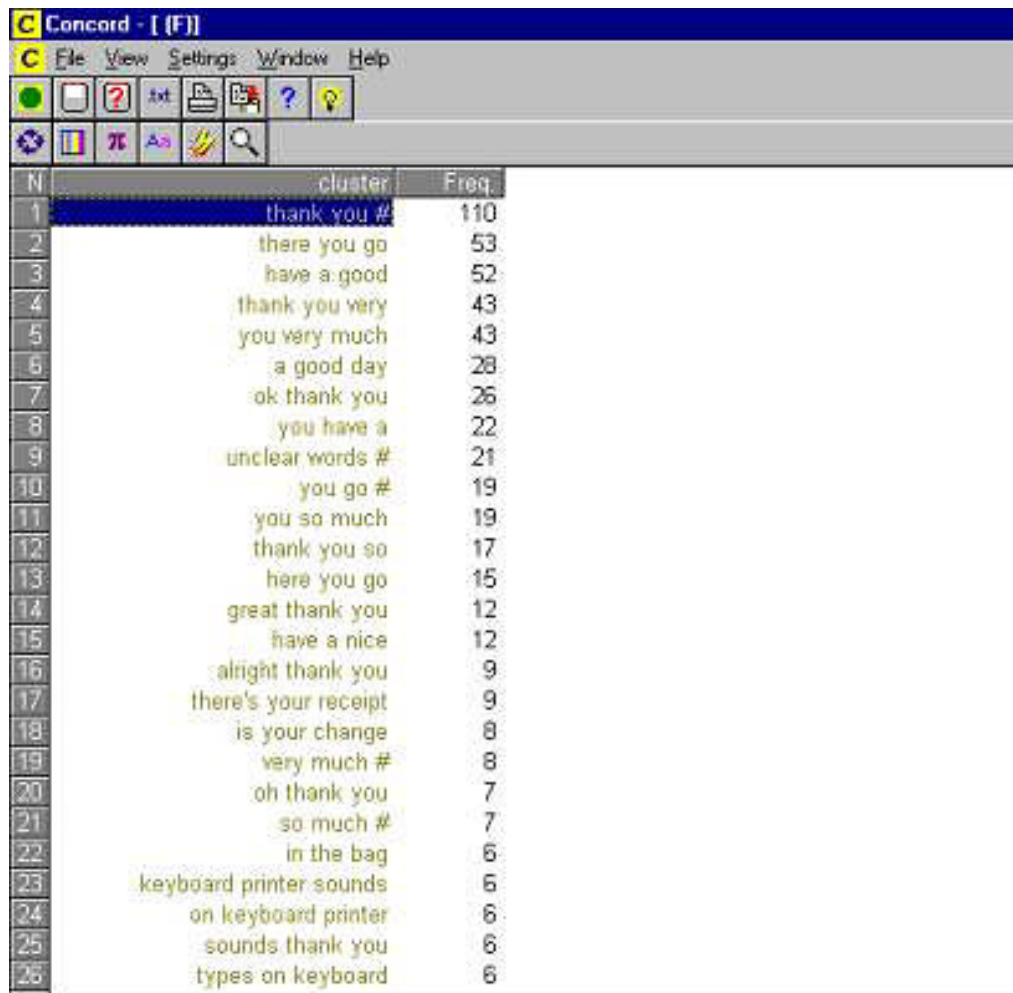


Figure 2. *WordSmith* plot distribution by text for the occurrence of *thank*

*WordSmith* also allows the user to compare word lists. The Key Word function allows the user to compare a given text to a target text or target register, which can be particularly useful for cross-register comparisons. For example, a teacher or researcher could compare biology textbooks to geology textbooks in order to see what lexical similarities or differences occur. The Key Word function provides a quick glimpse of what the text is about, since the list is not based on absolute frequency but rather the unique words that are frequent in the particular text.

The Cluster function is the *WordSmith* feature that is perhaps most innovative since it is quite powerful and can be very useful. With this function, the user can specify from two to eight word clusters from a concordance list and then see which words tend to co-occur (see [Figure 3](#)). Co-occurring words are often idioms or set phrases.



N	cluster	Freq.
1	thank you #	110
2	there you go	53
3	have a good	52
4	thank you very	43
5	you very much	43
6	a good day	28
7	ok thank you	26
8	you have a	22
9	unclear words #	21
10	you go #	19
11	you so much	19
12	thank you so	17
13	here you go	15
14	great thank you	12
15	have a nice	12
16	alright thank you	9
17	there's your receipt	9
18	is your change	8
19	very much #	8
20	oh thank you	7
21	so much #	7
22	in the bag	6
23	keyboard printer sounds	6
24	on keyboard printer	6
25	sounds thank you	6
26	types on keyboard	6

Figure 3. *WordSmith* screen with clusters

*WordSmith* also has a feature that allows the user to align two texts and create a new file that contains one displayed over the other. This is extremely useful for comparing translations or two versions of the same text. The texts are displayed in different colors for ease of reading. See Figure 4 for an example of this feature used to check a translation against the original text.

Original :	Der Knabe sagte diesen Gedanken dem Schwesterchen, und diese folgte. Allein such der Weg auf den Hals hinab war nicht zu finden. So klar die Sonne schien, ... (from Stifter's <i>Bergkristall</i> , translated by Harry Steinhauser, in <i>German Stories</i> , Bantam Books 1961)
Translation:	The boy communicated this thought to his sister and she followed him. But the road down the neck could not be found either. Though the sun shone clearly, ...
Aligned text:	
<G1>	Der Knabe sagte diesen Gedanken dem Schwesterchen, und diese folgte.
<E1>	The boy communicated this thought to his sister and she followed him.
<G2>	Allein auch der Weg auf den Hals hinab war nicht zu finden.
<E2>	But the road down the neck could not be found either.
<G3>	So klar die Sonne schien, ...
<E3>	Though the sun shone clearly, ...

Figure 4. Aligning two texts to check a translation (excerpt from *WordSmith* on-line manual)

The main advantage of *MonoConc Pro* over *WordSmith* is that it is much easier to use. For example, when *MonoConc Pro* is launched, a clear easy-to-use screen appears with a bar across the top, providing the options available. On the other hand, when *WordSmith* is launched there are many screens that appear, and until the user becomes familiar with the program, just getting the program going can be a bit of a challenge. For someone starting out with corpus analysis, and wanting to focus mostly on concordancing, *MonoConc Pro* is more user-friendly. The screens are clearer, and since they resemble the screens of many word processing programs, users may feel more comfortable.

In summary, both programs offer users powerful tools for searching texts and exploring how language is used in natural settings, thus providing valuable resources for teachers and researchers. However, the two programs have different strengths: for users who are less comfortable with computers, *MonoConc Pro's* interface is much more user-friendly than that of *WordSmith*. However, for those who are comfortable with computers and plan to carry out more powerful text analysis, *WordSmith* would be a better choice. So, while both *MonoConc Pro* and *WordSmith* offer attractive options for exploring texts, the best choice will depend on the specific goals and experience of the user.

---

### ABOUT THE REVIEWER

[Randi Reppen](#) is an Assistant Professor in Northern Arizona University's MA-TESL/PhD-Applied Linguistics Program, and the Director of the [Program in Intensive English](#). She is co-author of *Corpus Linguistics: Investigating Language Structure and Use* with Douglas Biber and Susan Conrad (1998). Her research interests include corpus linguistics and the use of corpora in materials development.

E-mail: [Randi.Reppen@NAU.EDU](mailto:Randi.Reppen@NAU.EDU)

### REFERENCE

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating language structure and use*. Cambridge, UK: Cambridge University Press.