Concordance and Diction

Comparative Review

Concordance
Platform: Windows NT 4.0 and 95/98
Requirements: 32MB RAM, 4MB disk space
Available from: http://www.rjcw.freeserve.co.uk/
Price: Online registration costs $89 (U.S. dollars) for a single copy. Second and subsequent copies cost $40 each + $10 handling fee for each online transaction. (Free trial download available.)

Diction
Platform: Windows 95 or higher
Requirements: 4MB RAM, 2.5MB disk space
Available from: http://www.scolari.co.uk/
Price: £80 (educational user discount) + £5 carriage for UK delivery. (Free trial download available.)

This review will address two software programs, Concordance and Diction 4.0. While both programs are essentially text analysis tools, their use and results are wholly disparate and do not lend themselves to direct comparison. As a program, Concordance conforms to the more stereotypical task of text analysis—it performs the eponymous duty upon either a full or select text. The end goal of Diction, however, is the investigation of semantic features based upon referencing the inputted text against a 10,000-word corpus.

Concordance

Concordance was created in 1999 by R.J.C. Watt, Senior Lecturer in English at Dundee University. It is important, when discussing a program like Concordance, to take into consideration the aspects of the software that differentiate it from the other programs that also provide concordancing capabilities. Many of the tools on the market offer similar results, in that you can build concordances, wordlists, basic statistics, etc, so rather than focusing too heavily on the expected traits, the review will focus more on the additional tools found within the product.

One of the first features to stand out is the software's ability to work on any size text extremely quickly. The software allows the user to create either a full or fast concordance—the full concordance being the more traditional processing of a full-text, with the fast option creating a concordance on the basis of user-selected words. As an added bonus, Concordance offers the translation of the results into a web version. This ability to save as HTML helps distinguish this software program from others of its type (examples of these results can be viewed at *The Web Concordances* home page, http://www.dundee.ac.uk/english/wics/wics.htm). Essentially the program is limited only by the user's computer. The more disk space and memory in the hardware the better the performance. The fast concordance option allows the user more control over the final result.
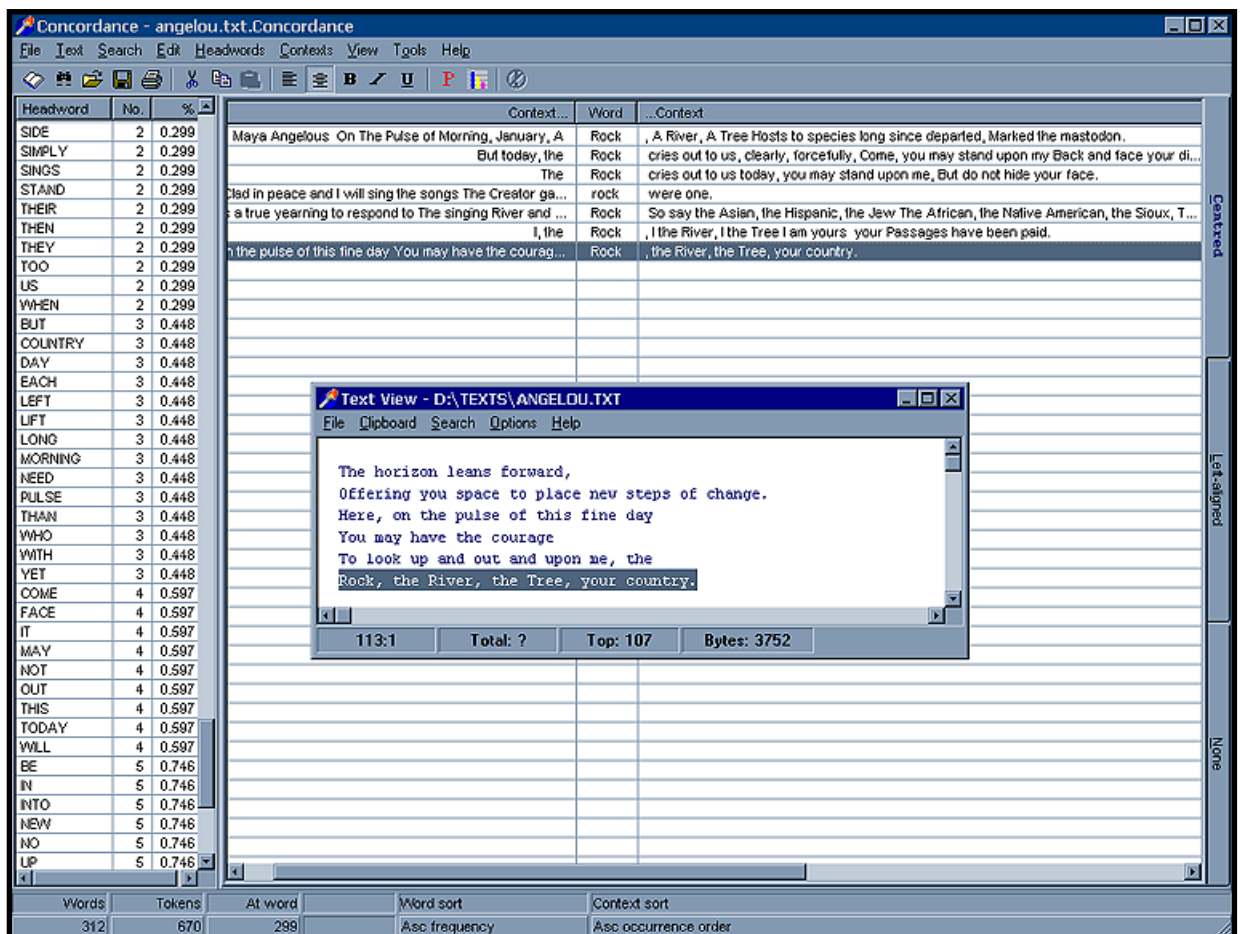

Fig.1. Concordance allows the user to view the wordlist, concordance and source text at the same time.
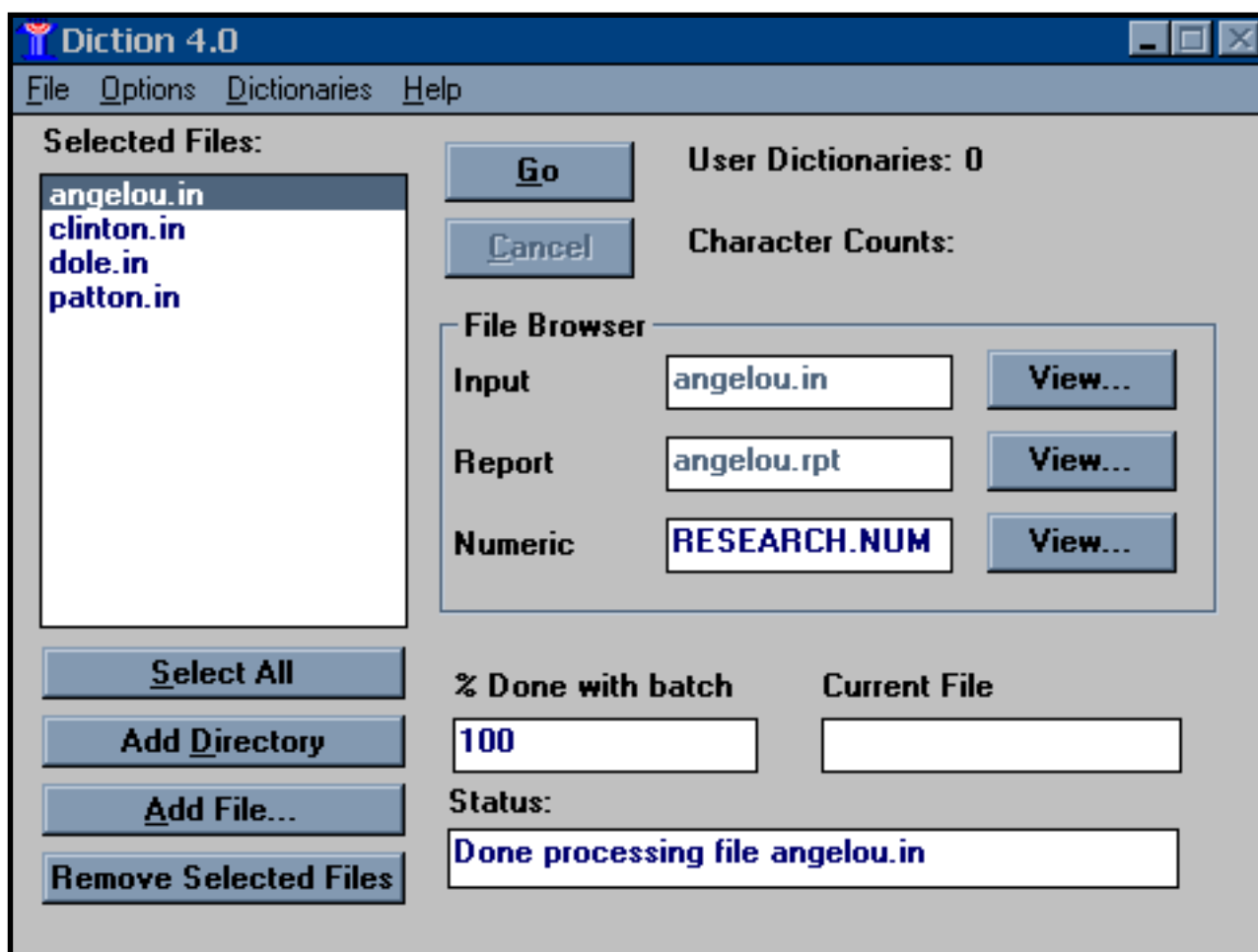
Fig.2 Diction 4.0's user-friendly minimalist interface.

The program is also very accommodating in terms of foreign languages and special characters. To analyze a foreign language text, it is only necessary to go to the Language Control button on the Tools menu and select the appropriate language. The software then adapts automatically to the inputted text. Concordance can also handle mixed language texts, as long as the languages have the same character sets. To work with special characters, the user can edit the recognized alphabet set to include any characters found in the text. Also, if there is text that needs to be ignored, the user can insert that text in a marker of their own choosing. Then by simply using the 'Text/Ignore' option and inputting the choice, the selections will be disregarded during the concordance process. The only disadvantage of this is that the 'Ignore' option is limited to five marker selections and does not accommodate an SGML markup system.

One of the optimal features of Concordance, and one that could easily be overlooked, is its extremely friendly interface. Once the program opens, the process of building a concordance is almost self-explanatory. And if there are questions, such as making changes concerning alphabet or language, then the Help function is quite thorough. The interface also allows the user to view the wordlist, concordance and source text at the same time (see Figure 1). This utility is enhanced by the fact that when the user chooses words in the wordlist the KWIC (Keyword in Context) changes automatically, and then if a line from the KWIC is chosen the text viewer immediately reflects that change. In other words, there is no need to toggle between screens or continually close windows before choosing

headwords—headwords, KWIC, and referenced source text can all be viewed simultaneously.

Diction

While Concordance prides itself on the fact that the software can process large quantities of text, Diction works primarily with political discourse, or rather small speeches of that ilk. The program processes imported plain text ASCII files, and recommends that texts are checked with a program such as Notepad to ensure that they conform to the requirement. The software is optimized for works that do not exceed 500 words in length. However, to compensate for this severe limitation, it can handle works with a maximum of 5,000 words. Unfortunately, as the 500 word length is required for standardization of results, any text that exceeds this size is broken down into smaller files to process. If a text is less than 500 words, it can be processed as Report Extrapolations, which will normalize the results through corrective counts.

In order to study the semantics of human discourse, Diction evaluates the text and then calculates scores based on five categories—Activity, Certainty, Commonality, Optimism, and Realism. Each category has a set of sub-categories, which are made up of lists of distinguishing terms, against which the text is compared. These dictionaries come with the software program and cannot be changed. However, if there are specific word lists that the user would like to include in the analysis, an option of including user-customized dictionaries is available. The limitation with this is that only ten custom dictionaries of 200 words each are allowed. The interface

(see Figure 2) is user-friendly in its minimalism. Simply select a file (or files) to process, click "Go" and the program outputs the results into the Report area of the File Browser. Diction is insular in that it produces results with little to no input from the user—it is not an interactive experience. To view the results of the process, simply click the View button and the report will open up in a Notepad screen. After establishing basic statistics about the text, the report provides a list of High Frequency Words and then moves on to the Dictionary Totals (see Figure 3).

Before using Diction to perform any kind of analysis a few things must be taken into consideration. First, this is not a tool created for largescale texts. It is not meant for use on novels or lengthy essays, although select analysis could be performed with user-customized dictionaries. The software is meant for the analysis of human discourse—fundamentally speeches of political origin. The 500-word text is clearly the optimal choice and while the software does allow for larger texts, it is very difficult to achieve consistent or holistic statistical output when the results are rendered into disparate files. Secondly, the dictionary that the texts are referenced against are derived from American political speeches. Obviously this will affect the results in terms of language analysis. Inputting a work created in the UK might prove an interesting study but the statistical output must be analyzed in light of the language differences. Again, this issue can be resolved with the inclusion of customized dictionaries. But users need to be aware of the situation before evaluating the results. Lastly, the users should take advantage of the fact that this is not a straightforward concordancing tool. While there are clear limitations with the software program (language, file size, little user input), Diction is quite specialized and provides a unique analysis of language. Its ability to examine language (esp. homographs) in terms of context and then make judgments based upon semantics allows the program a distinctive position amongst its counterparts.

Karen Wikander
Oxford Text Archive
karen.wikander@oucs.ox.ac.uk

angelou.rpt - Notepad

File Edit Search Help

DICTIONARY TOTALS

VARIABLE	FREQ.	PER CENT (FREQ/500)	NORMAL RANGE OF SCORES (#)	STANDARD- IZED SCORES	
Numerical Terms	2.00	0.40	0.30-- 15.04	-0.77	
Ambivalence	7.30	1.46	6.49-- 19.21	-0.87	
Self-reference	23.00	4.60	0.00-- 15.10	1.97	*
Tenacity	19.10	3.82	23.32-- 39.76	-1.51	*
Leveling	7.00	1.40	5.02-- 12.76	-0.49	
Collectives	5.00	1.00	4.04-- 14.46	-0.82	
Praise	3.84	0.77	2.77-- 9.59	-0.69	
Satisfaction	3.00	0.60	0.47-- 6.09	-0.10	
Inspiration	1.00	0.20	1.56-- 11.10	-1.12	*
Blame	3.00	0.60	0.06-- 4.16	0.43	
Hardship	7.00	1.40	1.26-- 10.48	0.25	
Aggression	6.00	1.20	1.07-- 9.79	0.13	
Accomplishment	6.50	1.30	4.96-- 23.78	-0.84	
Communication	7.71	1.54	2.21-- 11.79	0.15	
Cognitive Terms	4.54	0.91	4.43-- 14.27	-0.98	
Passivity	4.43	0.89	2.10-- 8.08	-0.22	
Spatial Awareness	12.50	2.50	4.17-- 19.85	0.06	
Familiarity	133.00	26.60	117.87--147.19	0.03	
Temporal Awareness	14.00	2.80	8.36-- 21.82	-0.16	
Present Concern	16.00	3.20	7.02-- 16.60	0.86	

Fig.3 Diction 4.0's list of High Frequency Words and Dictionary Totals.