Read me File for AntConc 3.1.2 (Windows and Linux)

##
Laurence Anthony, Ph.D.
Center for English Language Education in Science and Engineering
School of Science and Engineering
Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan
January 16th, 2006
##

AntConc started out as a relatively simple concordance program, but has been slowly progressing to become a rather useful text analysis tool. It is written in Perl 5.8 using ActiveState's excellent Komodo development environment. The program can be launched by simply double clicking on the .exe executable file, which can be downloaded from the Laurence Anthony Laboratory web site. The program can run under any windows environment including Win 98/Me/2000/NT and XP, and also Linux. If a user finds any problem launching the program under a particular OS, please let me know.

AntConc contains the following tools that will be explained separately.
Concordance
Concordance Plot
File View
Word Clusters
N-Grams (part of Word Clusters)
Collocates
Word List
Keyword List

Note that each tool can be accessed either by clicking on its 'tab' in the tool window, or using the function keys F1 to F7.

Concordance

The **Concordance** tool generates concordance lines (or KWIC: key word in context) lines from one or more target texts chosen by the user.

To produce a set of concordance lines of text, a user needs to perform the following actions:

1) Select one or more files for processing from using the 'Open File(s)' or 'Open Dir' options in the 'File Menu'. Files will be continually added to the system. The list of selected files is shown in the left frame of the main window.
2) Enter a search term on which to build concordance lines in the entry box on the left of the button bar.
3) Choose the number of text characters to be outputted on either side of the search term, using the increase and decrease buttons on the right of the button bar under the "Search Window Size" title. (default value is 50 characters)
4) Click on the 'Start' button to start the concordance lines results generation. (Note: The concordance generation can be halted at any time by clicking on the 'Stop' button)
5) Select a target word on which to rearrange the concordance lines, using the buttons to the right of the button bar. 0 is the search word, 1L, 2L... are words to the left of the target word, 1R, 2R .. are words to the right of the target word. Note that three levels of sort are possible, with the second and third levels not-activated when the software is first launched.
6) Click on the 'Sort' button to start the sorting process.
7) Move the cursor over the highlighted search term in one of the concordance lines. The cursor will change to a small hand icon. Clicking on the highlighted search term, will allow the user to view the search term hit as it appears in the original file via the **File View** tool (see below).

Search terms can be specified as being "words"(default) or "word fragments" by clicking on the search term option. Also, searches can be either case sensitive, or case insensitive (default). Searches can also be made using full regular expressions by choosing the "Regex" option. For details on how to use regular expressions, consult one of the many texts on the subject. E.g., Mastering Regular Expressions (O'Reilly & Associates Press) or type "regular expressions" in a web search engine to find many sites on the subject.

Note that the total number of concordance lines generated (hits) is shown in the top right of the AntConc window. This number will flash with the word "FINISHED" when processing has been completed, and will flash with the word "NO HITS", if not hits are generated for a particular search term. In this case, the concordance lines view will not be updated, and the previous set of concordance lines will remain visible.

Concordance Plot

Generating concordance plots can be achieved using the same actions as when using the **Concordance** tool. However, the **Concordance Plot** tool offers an alternative view of concordance lines. Here, all the hits for each file are plotted in the form of a 'barcode' indicating the position in the file where the hit occurred. The plot provides an easy way to see which files include the target search term, and can also be used to identify where the search term hits cluster together. An example of the use of the plot is in determining where specific content words appear in a technical paper, or when a character appears during the course of a novel or play.

The number of hits and length of each text is shown to the right of the barcode plot, and the plot itself can be enlarged or reduced in size using the zoom buttons.

If you move the cursor over the highlighted search term in one of the concordance lines. The cursor will change to a small hand icon. Clicking on the highlighted search term, will allow the user to view the search term hit as it appears in the original file via. the **File View** tool (see below).

File View

At any time a target file can be viewed in its original form using the **File View** tool.

To produce a view of the original file, a user needs to perform the following actions:

1) Select a file to view in the file list frame to the left of the main window.

2) If a search term has been specified, the search term hits will be highlighted throughout the text. Search options are the same as for the **Concordance** and **Concordance Plot** tools.

3) Use the "Previous Hit" and "Next Hit" buttons to jump to the appropriate hit in the file.

4) Change the search term and click on the 'Start' button to view other hits in the file.

Below is a list of Shortcuts unique to the **File View** tool.

SHIFT-Click = Jumps to the nearest hit in the window.
Use the 'Up' and 'Down' arrows on the keyboard to access the search term history.

Clicking on the highlighted text will generate a set of KWIC lines using the highlighted text as the search term.

Word Clusters

The **Word Clusters** tool is used to generate an ordered list of clusters that appear around a search term in the target files listed in the left frame of the main window.

The clusters can be ordered either by frequency or the start or end of the word. They can also be ordered by the probability of the first word in the cluster preceding the remaining words. All list orderings can also be inverted. Also, a user can select the minimum and maximum length (number of words) in each cluster, and the minimum frequency of clusters displayed. It is also possible to select if the search term always appears on the left or right of the cluster. (Note: In the current version, if more than one word is specified as the search term, only the first word will appear on the right, if the "Search Term on Right" option is selected.)

To produce a cluster list , a user needs to perform the following actions:

1) Choose the appropriate ordering options.

2) Press the 'Start' button.

At any time, the generation of the clusters list can be halted using the 'Stop' button.

Clicking on the cluster will generate a set of KWIC lines using the text as the search term.

A number of menu preferences are available with this tool. (See below).

N-Grams (part of Word Clusters)

The **N-Grams** tool is used to generate an ordered list of n-grams that appear in the target files listed in the left frame of the main window. N-grams are word n-grams, and therefore, large files will create huge numbers of n-grams. For example, n-grams of size 2 for the sentence "this is a pen", are 'this is', 'is a' and 'a pen'.

As with the **Word Clusters** tool, the n-grams can be ordered either by frequency or the start or end of the word. They can also be ordered by the probability of the first word in the cluster preceding the remaining words. All list orderings can also be inverted. Also, a user can select the minimum and maximum size (number of words) in each n-gram, and the minimum frequency of n-grams displayed.

To produce an N-gram list, a user needs to perform the following actions:
1) Click on the "N-Grams" option above the search entry box.

2) Choose the appropriate ordering options.

3) Press the 'Start' button.

At any time, the generation of the n-grams list can be halted using the 'Stop' button.

Clicking on the lexical bundle will generate a set of KWIC lines using the text as the search term.

A number of menu preferences are available with this tool. (See below).

Collocates

The **Collocates** tool is used to generate an ordered list of collocates that appear near a search term in the target files
 listed in the left frame of the main window.

The collocates can be ordered either by frequency, frequency on the left or right of the search term, or the start or end of

the word. They can also be ordered by the value of a statistical measure between the search term and the collocate. The value measures how 'related' the search term and the collocate are. Current possible statistical measures are listed below. All list orderings can also be inverted. Also, a user can select the span of words to the left and right of the search term in which to find collocates, and the minimum frequency of collocates displayed. If only a one-word span is required, for example, to see which words appear directly on the right of the search term, check the "Same" box, to keep the minimum and maximum span size the same.

Statistical Measures:
(MI) Mutual Information: Using equations described in M. Stubbs, Collocations and Semantic Profiles, Functions of Language 2, 1 (1995)
(T-Score) T-Score: Using equations described in M. Stubbs, Collocations and Semantic Profiles, Functions of Language 2, 1 (1995)

To produce a collocate list, a user needs to perform the following actions:

1) Choose the appropriate ordering options.

2) Press the 'Start' button.

At any time, the generation of the collocates list can be halted using the 'Stop' button.

Clicking on the collocates will generate a set of KWIC lines using the text as the search term.

A number of menu preferences are available with this tool. (See below).

Word List

The Word List feature is used to generate a list of ordered words that appear in the target files listed in the left frame of the main window.

The words can be ordered either by frequency or the start or end of the word, and the list can be inverted. The word list can also be generated in case-insensitive mode, where words in upper and lower case are treated the same (default) or case-sensitive, where words in upper and lower case are treated separately.

To produce a word list, a user needs to perform the following actions:

1) Choose the appropriate ordering options.

2) Press the 'Start' button.

At any time, the generation of the word list can be halted using the 'Stop' button.

A number of preferences are available when generating word lists, via the "Wordlist Preferences" option in the "Preferences" menu. (See below).

Clicking on the word will generate a set of KWIC lines using the text as the search term.

Different window panes can be either shown or hidden via the 'Word List Preferences' option in the 'Tool Preferences' menu

Keyword List

In addition to generating word lists, AntConc can compare the words that appear in the target files with the words that appear in a 'reference corpus' to generate a list of "Keywords" that are unusually frequent (or infrequent) in the target files.

To produce a keyword list, a user needs to perform the following actions:

1) Select a set of target files.

2) Go to the 'Preferences' menu and chose the 'Keyword Preferences' option.

3) Choose a statistical measure to assess the 'keyness' of the target file words. The default setting of Log Likelihood is recommended.

4) Choose a threshold for the number of keywords to be displayed.

5) Choose whether or not to view 'Negative Keywords' (target file words with an unusually low frequency compared with the frequency in the reference corpus)

6) Choose a reference corpus of text (.txt) files, in the same manner that the target files are chosen.

7) The reference corpus directory will be shown (if appropriate), and the list of reference corpus files will appear at the bottom of the Keyword Preferences option menu.

8) Click 'OK' in the Keyword Preferences menu, and return to the main Keywords window.

9) Choose suitable options for displaying the list of generated Keywords (in a similar manner to the options for generating a Word List).

10) Press the 'Start' button.

At any time, the generation of the keyword list can be halted using the 'Stop' button.

Clicking on the keyword will generate a set of KWIC lines using the text as the search term.

MENU OPTIONS
Menu options are divided into three groups, "File", "Global Settings" and "Tool Preferences". The option available in each category will be described below.

<FILE>
Options here relate to reading files into AntConc and writing files to the hard disk containing data of various types. There are also options to export all current settings to a file, and import user settings from a file. If a user settings file becomes corrupted for any reason, simply restart the program or use the "Restore Default Settings" option to return the program to its original state.

<GLOBAL SETTINGS>
Options here will have an effect on multiple tools in AntConc:

<File Settings>
In the File Settings menu, the user can choose to display the full path of a file or just the name. The user can also choose to show or hide any tags in the file. The tag boundaries can be specified.

<Tag Settings>

In the Tag Settings menu, the user can choose to display or hide any tags that are contained in the corpus files. If tags are to be hidden, the opening and closing tag markers must be specified. The default is <>.

<Wildcard Settings>
In the Wildcard Settings menu, uses have access to a wide range of wildcard characters which can be user defined. For example, to search for play, played or playing, use the following search term play*.

<Token Definition>
In the Token Definition menu, the user can choose which characters, numbers and so on will define a "word". For example, in some cases only letters will be considered words, but at other times, in might be desirable to include numbers, dashed and so on in the word definition. AntConc is fully Unicode compliant, meaning that it can handle data in any language, including all European languages and Asian languages. For this reason, the default options talk of letters in the broadest sense. Letters, for example, include all Japanese characters, if that language encoding is being used (see below). It is also possible for a user to define his or her own "word" definition.

<Highlight Color>
The color used for highlighting results can be changed using the menu option here.

<Level x Sort Color>
The color used for highlighting sorted results can be changed using the menu options here.

<Font Used for xxx>
The type, size and style of fonts used for displaying information can be changed using the menu options here.

<Font Encodings>
AntConc is fully Unicode compliant, meaning that it can handle data in any language, including all European languages and Asian languages. The language (encoding) of the data to be read by AntConc should be specified here. For example, f you are working with data saved in a Western language, it will usually be encoded in iso-8859-1 (default). On the other hand, Japanese texts are usually encoded in Shiftjis. By specifying the correct encoding, data from all languages can be processed correctly within AntConc. (Note: although the word 'Font' is used in this menu option, it really has nothing to do with 'fonts'. Encodings are used to define the mapping of characters).

<TOOL PREFERENCES>
Each tool (with the exception of **Concordance Plot** and **File View**) has a preferences menu, where settings can be fine tuned. All tool preference menus allow the user to show or hide the different frames in which the results are displayed. For example, the user can choose to hide the frame showing file names in the **Concordance** tool display window. Also, all tools have the option to arrange results either case sensitive or insensitive. If results are displayed case sensitively, words including capital letters will appear higher up in the lists.

<Concordance Preferences>
In addition to the above the following settings can be made:
Instead of arranging results by words to the left or right of the search term, it is possible to arrange the results by LETTERS to the left or right of the first letter of the search term. This makes it possible to search for spelling differences. The search term can also be chosen to be hidden in the KWIC lines, allowing instructors to quiz students on possible words to fit the gap. Note: at any time, if the <X> key is pressed while the results window has the focus, the search term can be hidden or shown.

<Word Clusters Preferences>
For this tool there are no additional settings that can be set other than those described above.

<Collocates Preferences>
In addition to the above the settings, the choice of statistical measure can be chosen here. Currently, two statistical

measure can be used: Mutual Information (MI) and T-Score. See above.

<Word List Preferences>
In addition to the above the following settings can be made:
The wordlist can be generated using all words, or a specific set of words, or ignoring a certain set of words (a stop list).
This is termed the "Wordlist Range". The range of words to be used (or ignored) can be entered directly by the user, or
can be stored in files which are then read by AntConc by pressing the 'Choose File' program. A combination of words in
 a file and words directly entered by the user can also be used.

A 'lemmas' list can be loaded from a file, which can then be used to generate a lemma list instead of a word list. When
the lemma list function is used, the 'lemma word form(s)' column will show the words in the corpus associated with
each lemma. Note that the 'lemmas' list should correspond with the token definition.

If you are using the lemma list provided by Yasumasa Someya (available on from the AntConc homepage), I would
suggest setting the token definition to include letters, and the 'dash' punctuation mark.

Note that words listed in the word list definition pane can be deleted or kept (while deleting all other files) using the
following shortcuts.
SHIFT-click = Selects continuous lines
CTRL-click = Selects discontinuous lines
DELETE = Deletes any selected lines that span across all window panes
INSERT = Keeps any selected lines and deletes all others

<Keyword List Preferences>
In addition to the above the following settings can be made:
As described in the section on the **Keyword List** tool, to generate a keyword list, the user needs to specify a
reference corpus, and a statistical measure of 'keyness'. Although, the default options for the 'keyness' measure and
threshold values are recommended, changes can be made in this menu.

SHORTCUTS

Here is a list of Shortcuts that apply to all tools using window panes for results.

CTRL-C = Copies the currently selected text
CTRL-A = Selects all text in the window pane
ALT-A = Selects all text in all window panes showing

Double click = Selects the current word
Triple click = Selects the current line in the window pane
SHIFT-click = Selects continuous lines across all window panes showing
CTRL-click = Selects discontinuous lines across all window panes showing

DELETE = This deletes any selected lines that span across all window panes
INSERT = This keeps any selected lines that span across all window panes, and deletes all others

Use the 'UP' and 'DOWN' arrows on the keyboard to access the search term history.

Results can be either saved to the clipboard, saved to a text file (.txt) or saved to a new window using keyboard
commands, the appropriate option in the 'File Menu', or by clicking on the "Save xxx Window" button in each tool,
respectively.

Other Comments

All new editions and bug fixes are listed in the revision history below. However, if a user finds a bug in the program, or has any suggestions for improving the program, please let me know and I will try to address the issues in an future version. Indeed, the revisions that have been made are largely due to the comments of users around the world, for which I am very grateful.

This software is available as 'freeware' (see Legal Matter below), but it is important for my funding to hear about any successes that people have with the software. Therefore, if you find the software useful, please send me an e-mail briefly describing how it is being used.

Acknowledgments

I would like to say thank you to the users of AntConc who have taken the trouble to e-mail me with feedback on the software and suggestions for improvements and/or changes.
The development of AntConc is supported by a Grant-in-aid for Scientific Research by the Japan Society for the Promotion of Education, Science, Sports and Culture, Japan (No. 16700573), and by a Waseda University Grant for Special Research Projects, Japan (No. 2004B-861).

Legal Matter

AntConc can be used freely for individual use for non-profit research purposes, and freely distributed on the condition that this read me file is attached in an unaltered state. If the software is planned to be used in a group environment, please let me know how you plan to use the software, and I will then determine if you can have permission to use it. The software comes on an 'as is' basis, and the author will accept no liability for any damage that may result from using the software.

Known Issues

1. On Linux systems, with the "Open File(s)" option, there appears to be a maximum number of files that can be opened at a single time. If more than this number is selected, none will be read into AntConc. On my computer, the maximum number appears to be around 950. To get around this problem, I advise the user to either select files in two or three batches, or use the "Open Directory" option, for which there is no maximum limit. On WINDOWS, a different file open method is used, but this creates ghost dialog boxes if the main File Open Dialog box is moved. Another strange bug that I cannot fix.

2. When a large number of concordance lines are generated (or words or keywords), the scroll bar becomes sensitive to where on the bar the user clicks and drags to view lower down entries. Sometimes this results in a user not being able to view the last lines unless the cursor is repositioned on the scrollbars. The is an annoying bug in the scroll bar subroutine (not mine!) and I am waiting for someone to fix this.

3. In the **Word Clusters** tool, if more than one word is specified as the search term, only the first word will appear on the right, if the "Search Term on Right" option is selected.

4. There is a strange but serious bug that causes the program to crash if the 'hide tags option' is activated, and then the File View tool is used before using any other tool. I do not understand the cause of the crash, but a solution is being investigated now. Also, the bug seems to only appear on certain machines at certain times. Possibly this is a problem related to using the program with files stored on the Desktop under a Japanese version of Windows. Therefore, I advise that users do not store files under non-Latin1 (Western) path names, for example, the 'Desktop' on Asian systems.

Revision History

3.1.2

This is a semi-major upgrade containing a new Lemma List tool, and numerous bug fixes and interface improvements.
A list of all major changes is below:
Binding to plot canvas lines to allow jumping to hit in file (same as in concordance view).
Binding to file list allowing the user to simply click on the file name when using the File View tool to view the file
History feature for all entries (use up and down arrows on keyboard)
Much faster concordancer processing (especially sorting... up to 10 times faster)
Ability to clear content of tools (Clear Tool, Clear All Tools, Clear All Tools and Files)
Ability to save list of all loaded files in settings file
Redesigned "one or more words" and "any one word" wildcards to act more sensibly. Now, the wildcards incorporate
any non-tokens between possible words. Therefore, use "is#dog" and "is@@dog" etc. to search for the hit "is a dog" in
the sentence "This is a dog."
New Lemma tool allowing all lemmas of a word list word to be displayed. Note: A lemma list is required.
Added bindings to Wordlist Range list to allow words to be deleted (<Delete> button) or kept (<Insert> button) as with
other tools.
Swap button to switch the main and reference corpora when doing keyword analyses. (Accessible from the Keyword
preferences Menu).
Fixed serious bug that caused the program to freeze or crash when a incorrectly formed regular expression was used as a
 search term.
Fixed serious bug that caused some word, keyword and font encodings to not be loaded correctly from the user settings
files
Fixed bug that caused the 'Reset' buttons to not work correctly.
Fixed bug that caused 'user defined token definitions' to not be loaded correctly.

3.1.1

This is a minor upgrade although it contains a new T-Score stats feature, and has a new editing feature, allowing you to
select and then delete, or keep certain results lines. A list of all major changes is below:
1) Added T-Score to stats measure in the Collocates Tool.
2) Changes the Collocates Tool menu to allow one a several statistical measure to be chosen.
3) Adding a feature so that any selected results lines can be either deleted or kept (after deleting all others).
4) The tag settings options in the File Settings menu have now been moved to a separate menu.
6) Labels in the software (in particular those related to language encodings) have been slightly altered to make them
clearer.
7) Improved the warnings given when the Collocates Tool is not in sync with the Word List Tool, during the calculation
 of collocate significance.
8) Changed the internal workings to allow LINUX and WINDOWS ports of AntConc to be easily created from the same
 source code. This should allow future versions to be released at the same time for both operating systems.
9) Fixed a bug that caused the Transition probabilities calculation to not initialize correctly. This meant that although the
 first calculation was correct, all later calculations would be give false measurements.

3.1

AntConc 3.1 is a major, major upgrade. I was very tempted to name this AntConc 4.0, but in the end chose to keep it in
the 3s. To list the improvements and changes will take up many pages, so only the major differences are listed here in
no particular order. Of course, with so many changes, there will inevitably be new bugs that have crept into the
program. However, I hope you will find that this release is an improvement over previous releases.
1) Implementation of new **Collocates** tool.
2) Total reworking of the programming underneath all tools for performance enhancements.
3) Adding of "sort" features to many tools, enabling quick and efficient re-ordering of results information.
4) Sensible treatment of word case in all tools. Now searches are truly case sensitive or insensitive, and results can be
displayed according the
case setting.
5) All settings can now be imported or exported to a file. This allows users to easily upload their preferred settings,
avoiding the need to constantly make changes to the defaults. This is a HUGE improvement (I think!).

6) The "Open Dir" option has been restored. File types that will be read into AntConc via the "Open Dir" option can now be specified in the "File Settings" menu.

7) AntConc is now FULLY Unicode compliant. It should work with any language in the world. I am very interested to hear how AntConc works in different languages. Please let me know. Token definitions can be made using all Unicode character classes, or a user defined token definition can be made. However, I would advise against using the user defined token definition, as it is so easy to overlook possible characters that might need to be processed.

8) As AntConc is now fully Unicode compliant, all possible encodings of characters are listed under the Encodings menu option. For English and all other Western Language, the default option, iso-8859-1 should work fine. Note that many Windows systems actually save files in the cp-1252 encoding which resembles iso-8859-1 but is a little different, just to confuse people! Users with Japanese texts will probably find shiftjis to be the best option.

9) Some users have said they could not open the font selection windows. This problem has been fixed.

10) Many new ordering options have been added, for example, from word ends.

11) The overall design of the interface has been improved, allowing for new options to be easily added without cluttering the display. Also most menu options now use simple checkboxes instead of the confusing "yes", "no" radio boxes used in previous versions.

12) Some of the tool names have been changed. In particular "lexical bundles" are now called "N-Grams", which is a more familiar term.

13) The Word Stems feature has been removed. This only worked with English texts, and the results were of questionable value. If a user would like to see the feature in a future version, please let me know.

14) The size of the program has increased by around 1MB. This is so that all the fonts and encodings for true internationalization of AntConc can be included.

15) This read me file has been largely rewritten. I hope there are fewer spelling mistakes this time!

16) Many, many small bugs have been fixed, mainly related to the way the interface responded to user actions.

3.0.1
A few bugs in 3.0 have been corrected. Also, the View Files tool has been redesigned to work much, much faster. To do this, the program now does no processing at the end of lines. Therefore, ambiguous line endings will stay ambiguous!
1) Improved performance of View Files tool
2) Corrected bug which caused the first file in Keywords to appear even after the list of files is deleted from the system.
3) Added feature to allow files to be added to the system from different folders. Now, if a file is opened into the system it will be added to the current list without deleting the previous list files. The same applied to files added in the Keywords List preference menu.

3.0
There have been so many changes that it is almost impossible to list them. Here are some of the more obvious differences between AntConc3.0 and previous versions.
1) Wildcard implementation
2) Word stem implementation
3) Save current window feature added to all tools (except 'Concordance Search Term Plot')
4) Three levels of sort implemented
5) Complete redesign and implementation of the 'View Files' tool, making it generate results much faster. (But it is still too slow).
6) Hyperlinks added to the results of all tools.
7) Rearrangement of the menu options
8) Redesign of the file list window, allowing one or more files to be closed.
9) Redesign of the main results windows, placing information in different window 'panes'. Each pane can be resized, or hidden.
10) Redesign of the data selection methods. The selection methods now comply with most other software products.
11) Complete redesign of many sub-routines enabling quicker processing.
12) Many, many bugs found and corrected. Please tell me if you find a bug because I WILL correct it.

2.6.0
1) Corrected a bug (introduced in version 2.5.3) which caused cluster lists to be not be displayed alphabetically.

2) Corrected a bug (introduced in version 2.5.3) that prevented the 2nd sort color to be properly updated from the settings menu. Thanks to a user for pointing this out to me.

3) Corrected a bug (introduced in version 2.5.3) that prevented files on a Japanese system to be open correctly. This bug is related to the new implementation of the Perl programming language. Please point out any bugs due to the new version of Perl, as they are very difficult for one person to discover.

4) Added an option to either show or ignore tags embedded in files (which is useful when processing HTML or XML files).

5) Added a 'lexical bundles' option to the Word Clusters tool, that generates "word n-grams".

6) Deleted the "Open Directory" option, as it prevented files other than .txt from being uploaded.

7) Added menu setting options to allow the font for searches and results to be changed.

8) Added menu setting options to show Japanese fonts (in shift_jis encoding) to be displayed properly.

9) Reduced the overall size of the program by approx. 1 MB.

10) A number of small bugs were corrected.

2.5.4
1) Corrected a bug (introduced in version 2.5.3) which caused word lists and keyword lists to be not be displayed alphabetically.

2.5.3
1) Adjusted the positioning and size of application widgets so that the application will be displayed properly on a 800 x 600 or higher resolution monitor.

2) Corrected several spelling errors in this read me file.

3) Adjusted the parameter settings for the Word Clusters tool.

4) Re-labeled several buttons so that they display properly on low resolution monitors.

2.5.2
1) Enabled a new 'Clusters' tool, for generating word clusters centered on a target search term.

2) Re-ordered Open File and Open Dir options in the File Menu. I think file navigation combined with 'select all files' <CTRL A> is easier than directory navigation,

3) Fixed a couple of small, insignificant bugs.

4) Fixed some spelling errors in this read me file.

2.5.1
1) Enabled a new feature, whereby clicking on the search term hit in any concordance line, allows the user to view the hit in the original file via the View Files tool.

2) Fixed a small bug in the View Files tool, which caused the searches to ignore the 'Case' setting.

3) Improved the performance when generating View Files by caching already processed files

2.5.0
A fairly major upgrade since 2.4.1
Here is a list of changes that have been made

1) Bug fix. When viewing files, and locating the next or previous hit, if the target file was changed and the hit number did not exist in the new file, the program would crash. This problem has been fixed.

2) Extension: In the view file window, hits would only appear if they occurred on a single line in the original file. This would result in different numbers of hits depending on if the search was made in the concordance window or the view files window. The view file processing has been completely revised enabling view file searching to correspond exactly with that used in the concordance window. Unfortunately, this has resulted in a small loss in performance when generating the highlighting in the view file. Also, clicking in the View File window now allows the user to immediately jump to the nearest hit.

3) The ability to show or not show full path names to files has been added as a system preference

4) The ability to show or not show file names in the concordance window has been added as a system preference.

5) The ability to set a wordlist 'range' has been added as a system preference.

6) Highlighting in the View Files tool has been changed to make the hits easier to see.

7) Pop-up windows that showed how many concordance hits were generated, and that reported when no hits were found have been removed. Instead, the status of the concordance hit processing is now shown in the top right of the main window.

8) Many small bugs relating to how the various tool displays are updated after preference changes are made have been corrected.

9) Processing that blocks user events (such as mouse clicks etc.) have been reduced.

10) The internal workings of the program have been re-written so that problems and future additions can be easily handled.

11) The general layout of the README file has be re-designed.

2.4.1
New since AntConc2.4.1 is the ability to choose whether or not to view 'Negative Keywords'. These are words in the target file that have an unusually 'low' frequency. In previous versions of AntConc, Negative Keywords were not distinguished from Keywords. However, now they are treated separately, and if the user chooses to display them, they appear after the Keywords, with a highlight color.

2.4.0
A major upgrade since 2.3.0
First, progress indicators were added to 'pages' of AntConc. Second, a new file view feature was added to view target files in their original state. Third, a keyword generation feature has been added using log-likelihood and chi-squared methods. Finally several bugs were found, in particular, bugs centered around the wordlist generation feature. This feature of the software should work much quicker now. Also, the user can interrupt the processing of files in any 'page' of the software.

2.3.0
A major upgrade since 2.2.3
First, the ability to view concordance search results as a barcode plot graph and a feature to produce wordlist according to different criteria were added. Numerous bugs centered around the way the software entered a 'Busy' mode were corrected. The main core of the software was also updated resulting in a quicker, 'cleaner' processing of the data. Performance improvements should be noticed as a result.

2.2.3
Updated file and directory selection dialog boxes to run smoothly in a Windows environment. Also, changed the default colors for sort highlighting, and search window frame size. A number of small bugs were also corrected

2.2.2
Corrected critical fault with compiler than caused program to expire when evaluation version of ActiveState Perl Development Kit expired. Sorry folks!! I didn't realize this would happen!!

2.2.1
Corrected bug which prevented new concordance lines being generated if the search term was left the same and then new files were selected. Port to Linux also completed.

2.2
Designed new subroutines for selecting directories and files to solve rendering of dialog windows problems. This also enables an easier port to Linux.

2.1
Added a second level of sort. Added ability to restrict searches to full-words only, case sensitive. Added ability to search using full Perl implemented regular expressions. Added ability to save results either to a file or the clipboard

2.0
Added new sort feature, for rearranging concordance lines. Tidied up the interface. Made the system more robust for novice users. (Now bad input will not cause the system to crash so easily).

1.1
Added binding to allow return key to launch concordance search. Also, recompiled software so that no console is required.

1.0
First version

Laurence Anthony