AntConc Manual

L. Xu P. Chen F. Liu Z. Zhang Y. Zhang

Jun 13, 2018

Contents

1 Introduction **1**

2 Tools **3**
- 2.1 Concordance . . . 3
- 2.2 Concordance Plot . . . 7
- 2.3 File View . . . 7
- 2.4 Clusters . . . 8
- 2.5 Collocates . . . 10
- 2.6 Word List . . . 11
- 2.7 Keyword List . . . 12

3 FAQ **21**
- 3.1 Facts about "Antconc" . . . 21
- 3.2 About Users . . . 21
- 3.3 Performance Issues . . . 22
- 3.4 About Shortcuts . . . 22
- 3.5 Saving Results . . . 22
- 3.6 Comments/Suggestions/Bug Fixes . . . 22

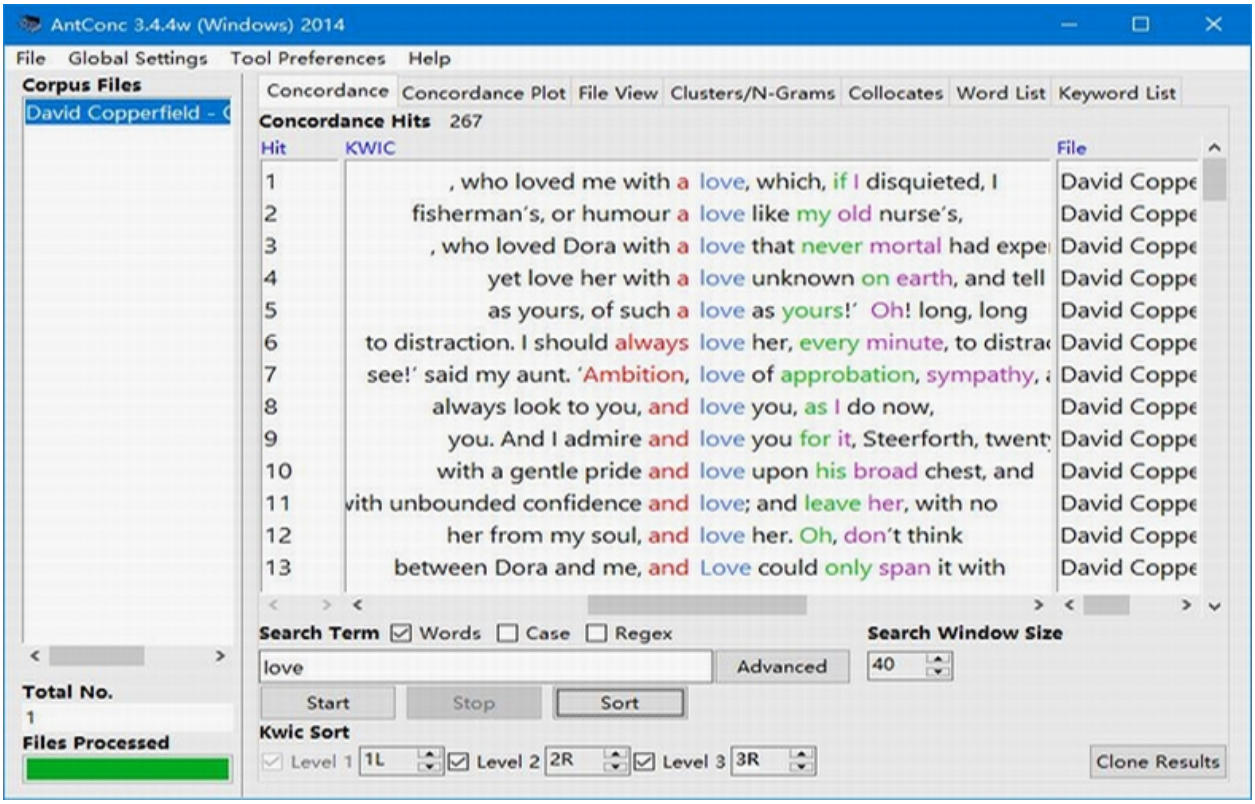
4 References **23**

CHAPTER 1

Introduction

AntConc is a freeware corpus analysis toolkit for concordancing and text analysis.

It runs on any computer running Microsoft Windows (tested on Win 98/Me/2000/NT, XP, Vista, Win 7), Macintosh OS X (tested on 10.4.x, 10.5.x, 10.6.x), and Linux (tested on Ubuntu 10, Linux Mint). It is developed in Perl using various compilers to generate executables for the different operating systems.



CHAPTER 2

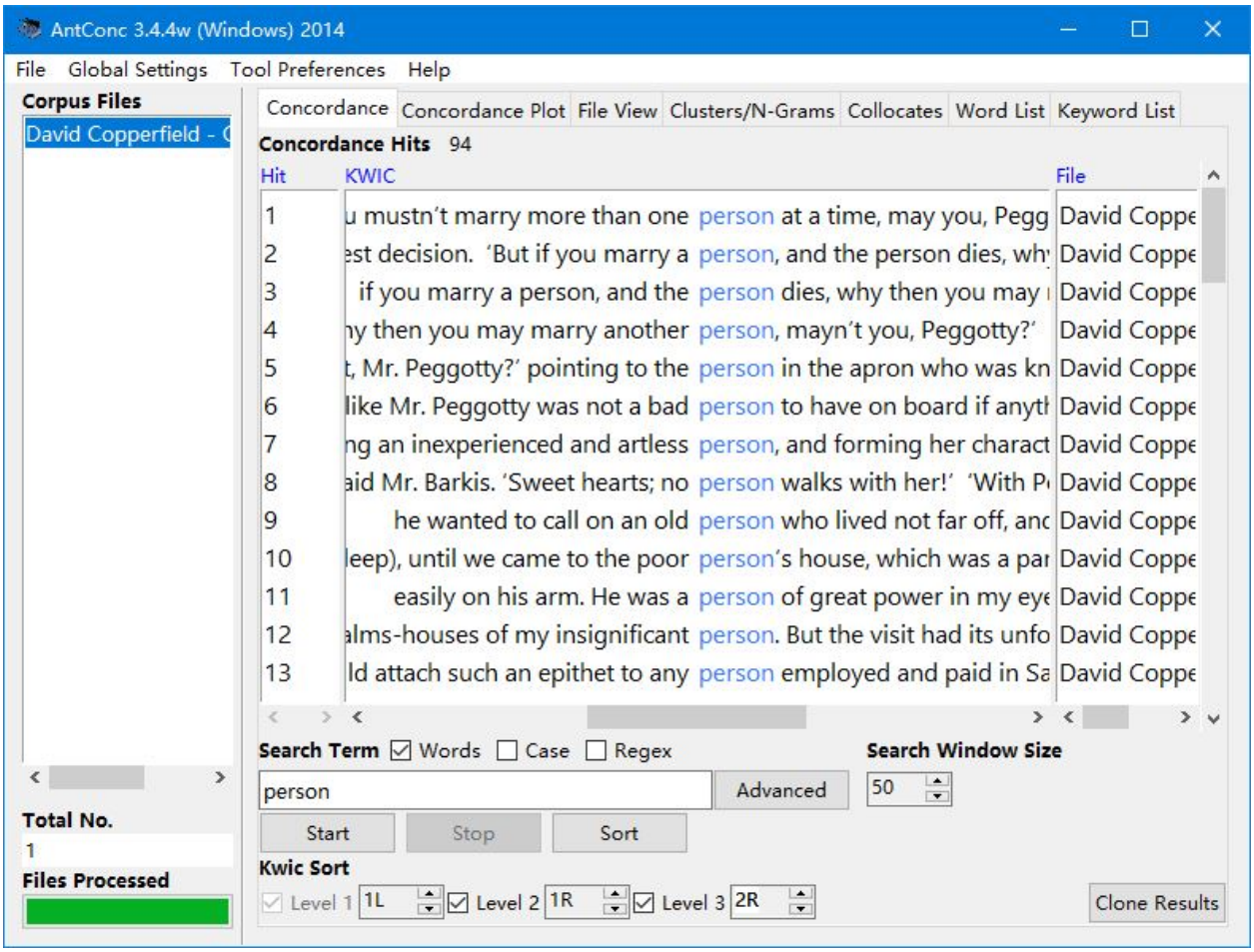
Tools

2.1 Concordance

This tool shows search results in a 'KWIC' (KeyWord In Context) format. This allows you to see how words and phrases are commonly used in a corpus of texts.

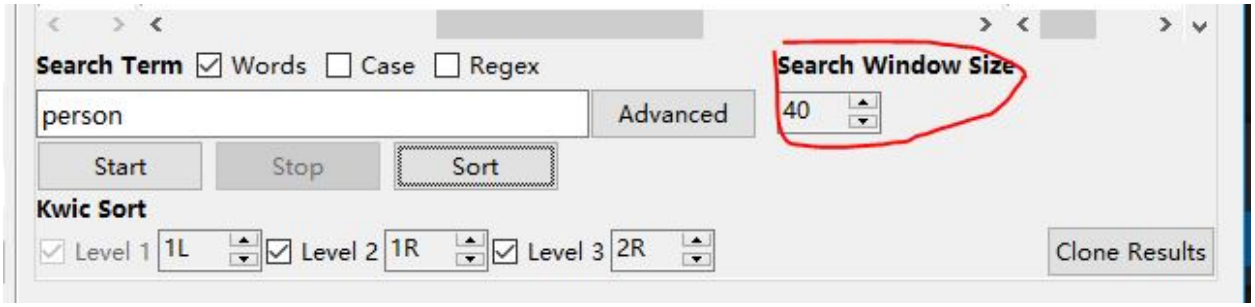
The following steps produce a set of concordance lines from a corpus and demonstrate the main features of this tool.

1. Select one or more files for processing from using the 'Open File(s)...' or 'Open Dir...' options in the 'File' menu. 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 search box.
3. Click on the 'Start' button to start the concordance lines results generation. The concordance generation can be halted at any time by clicking on the 'Stop' button.

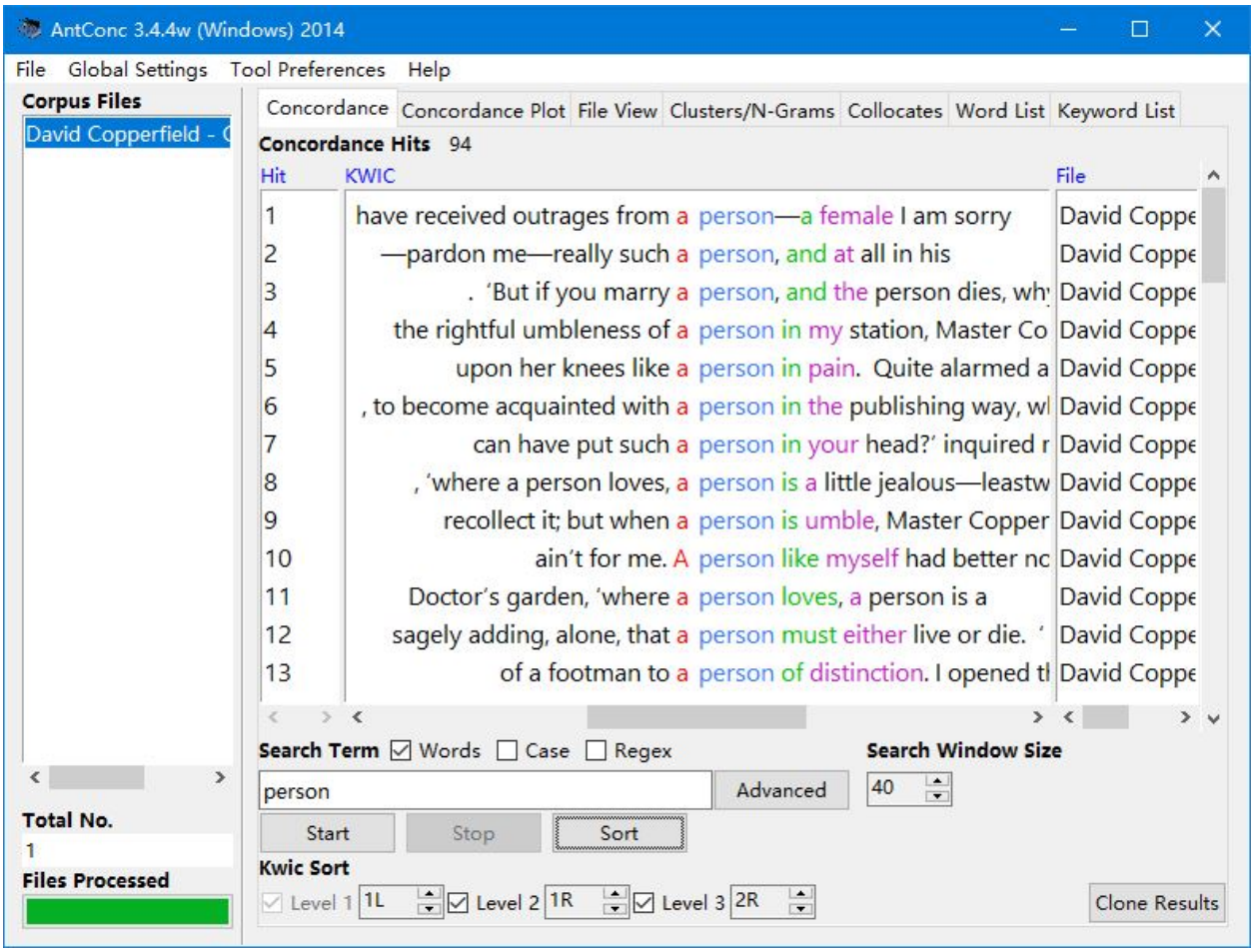


There are additional actions available.

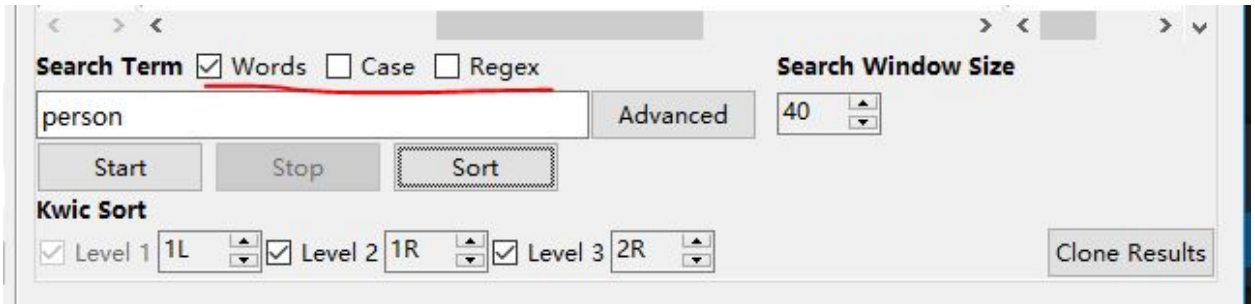
- You can choose the number of text characters to be outputted on either side of the search term, using the increase and decrease buttons on the right of the button bar under the "Search Window Size" title. (default value is 50 characters)



- Use the Kwic Sort options to rearrange the concordance lines at three different levels. 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.



- 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 you to view the search term hit as it appears in the original file via the File View Tool (see below).
- You can use the "words" checkbox above the search term inputbox to search words or characters, the "case" checkbox to search in case insensitive or sensitive mode and the "regex" checkbox to use regular expressions. AntConc supports Perl regular expressions. You can find more infomation here: Regular Expressions Quick Start

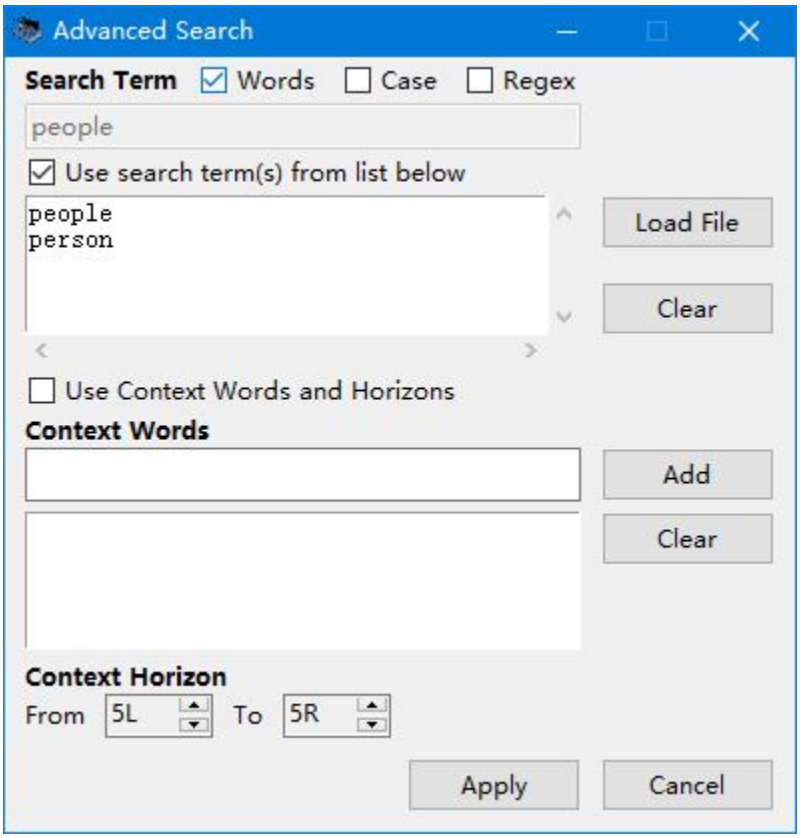


- Click on the "Clone Results" button to create a copy of the results so that different sets of results can be compared.

2.1.1 Advanced Options

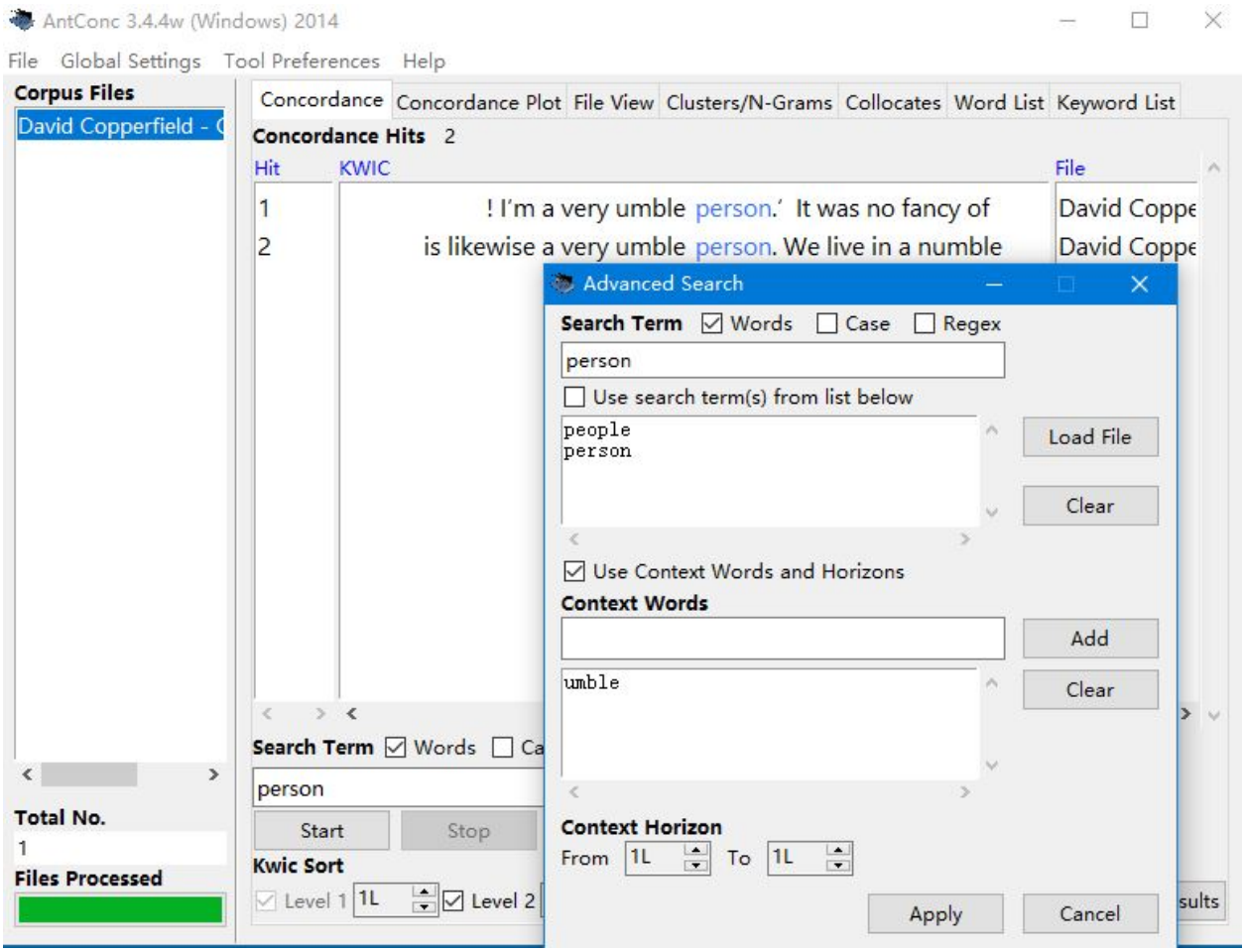
Click the "Advanced" to use advanced options.

- You can search multiple words by typing or by loading in a list of search terms from a file.



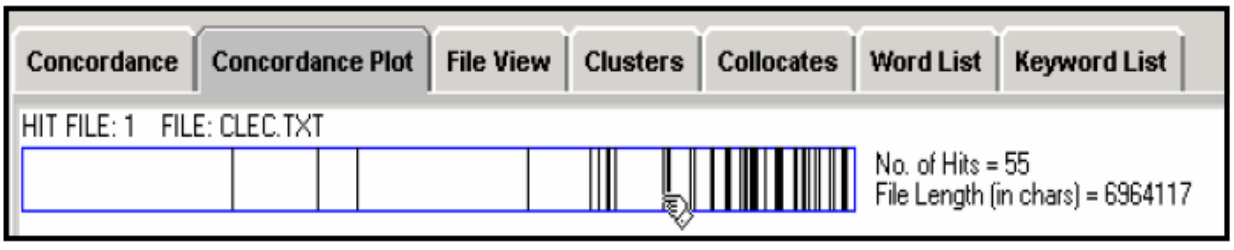
- The second advanced search option allows you to define context words and a context window within which the search term(s) must appear.

For example, to search for "person" where the word "umble" is the ajacent word in the left, set the search term as "person," the context word as "umble," and set the context window as 'From' 1L 'To' 1L.



2.2 Concordance Plot

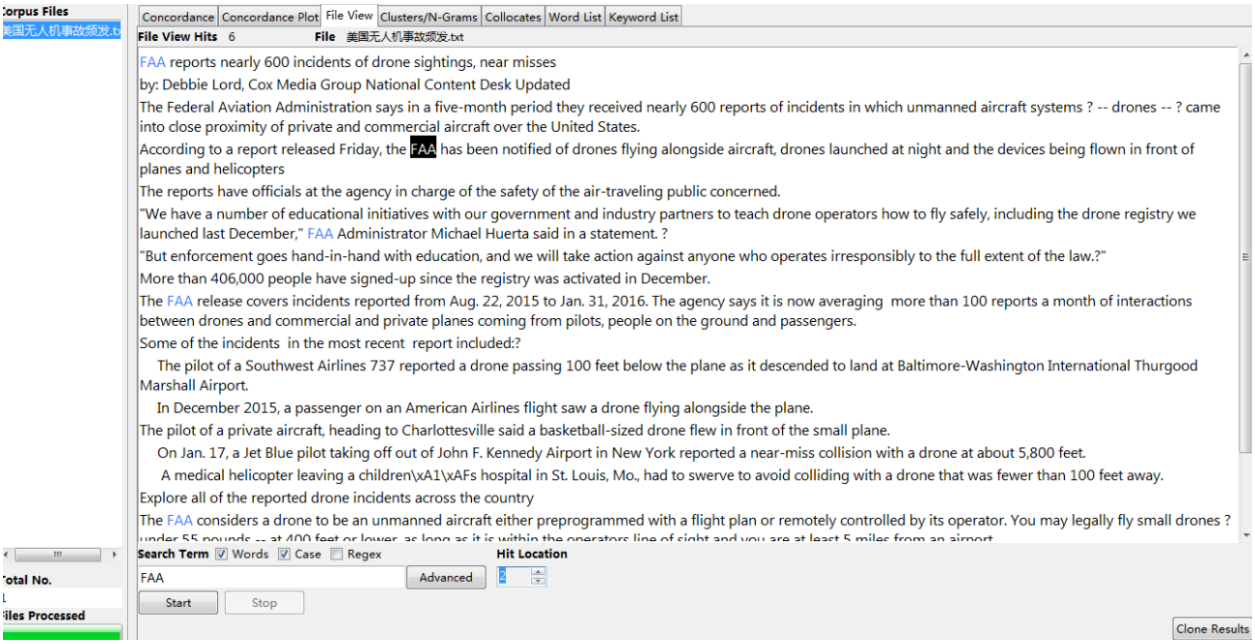
This tool shows concordance search results plotted in a 'barcode' format, with the length of the text normalized to the width of the bar and each hit shown as a vertical line within the bar. This allows you to see the position where search results appear in target texts.



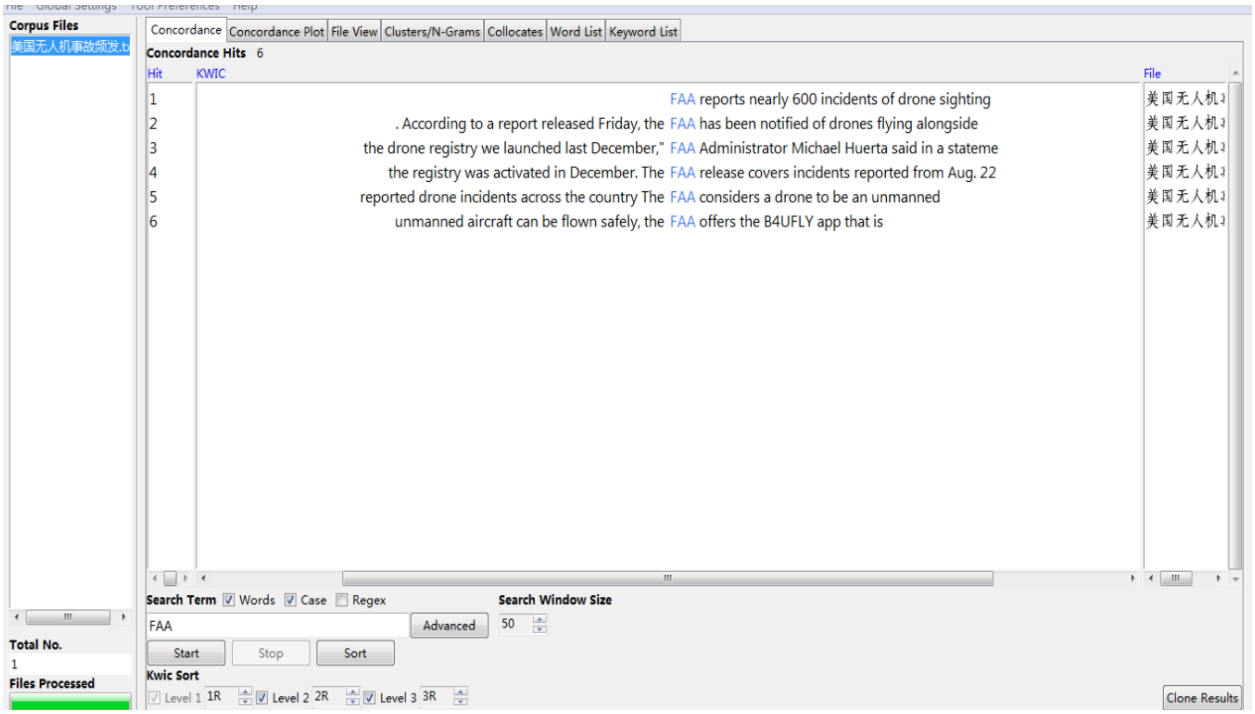
2.3 File View

This tool shows the raw text of individual files. This allows you to investigate in more detail the results generated in other tools of AntConc.

For example, if you search “FAA”, it will be highlighted throughout the text. And you can see the context of“FAA”. Search options are the same as for the Concordance Tool and Concordance Plot Tool.



If you click on the searched term “FAA”, you will get a set of KWIC lines.



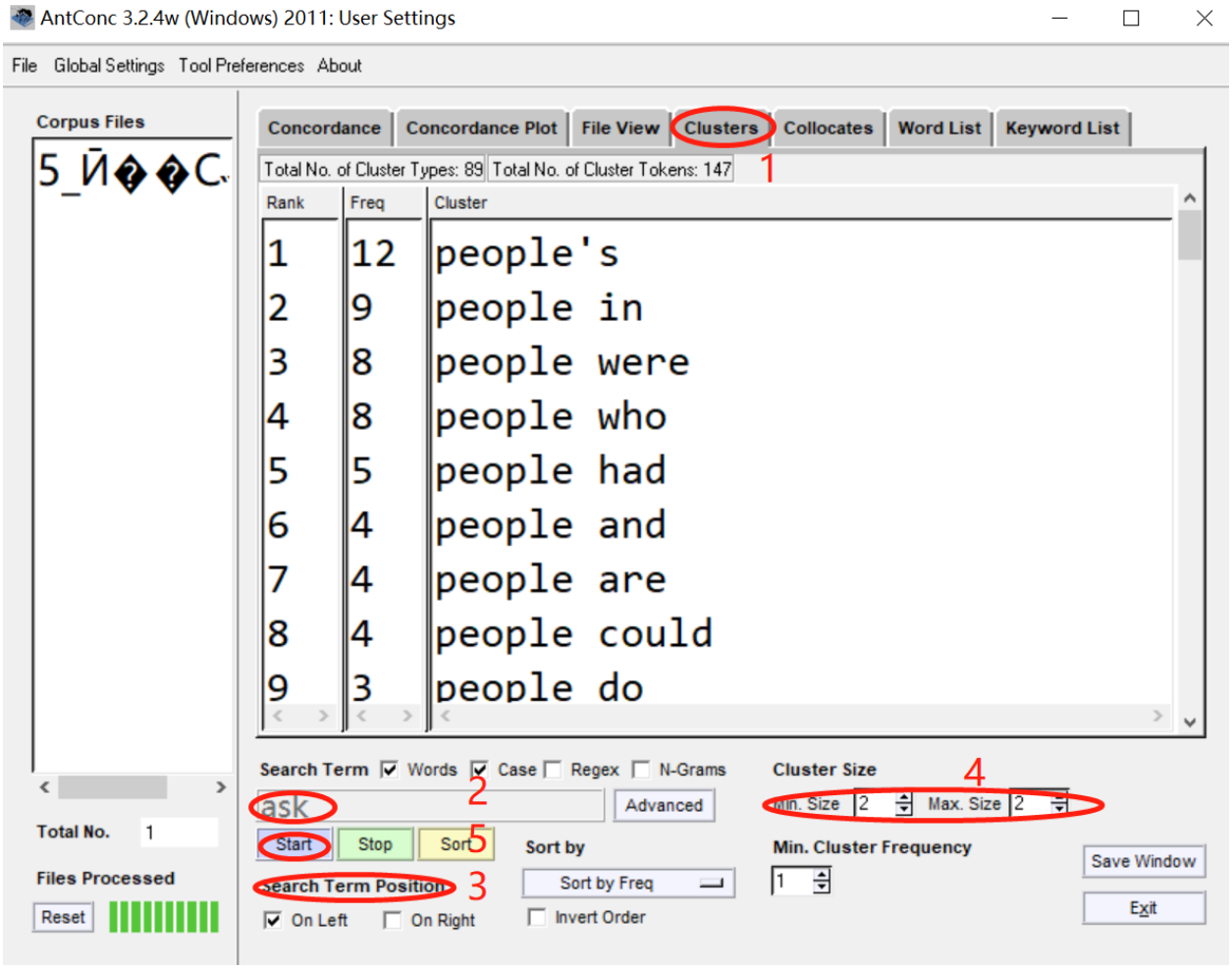
2.4 Clusters

This allows you to search for a word or pattern and group (cluster) the results together with the words immediately to the left or right of the search term.The clusters can be ordered by frequency, the start or end of the word, the range of

the cluster (number of files in which the cluster appears), or the probability of the first word in the cluster preceding the remaining words.

The following steps produce a set of clusters from a corpus and demonstrate the main features of this tool.

1. Click on 'Clusters'.
2. Enter a search term, like'ask'.
3. Select a 'Search Term Position'.
4. Select a 'Cluster Size'.
5. Click on 'Start' for a search.



In addition to the above, the following settings can be made:

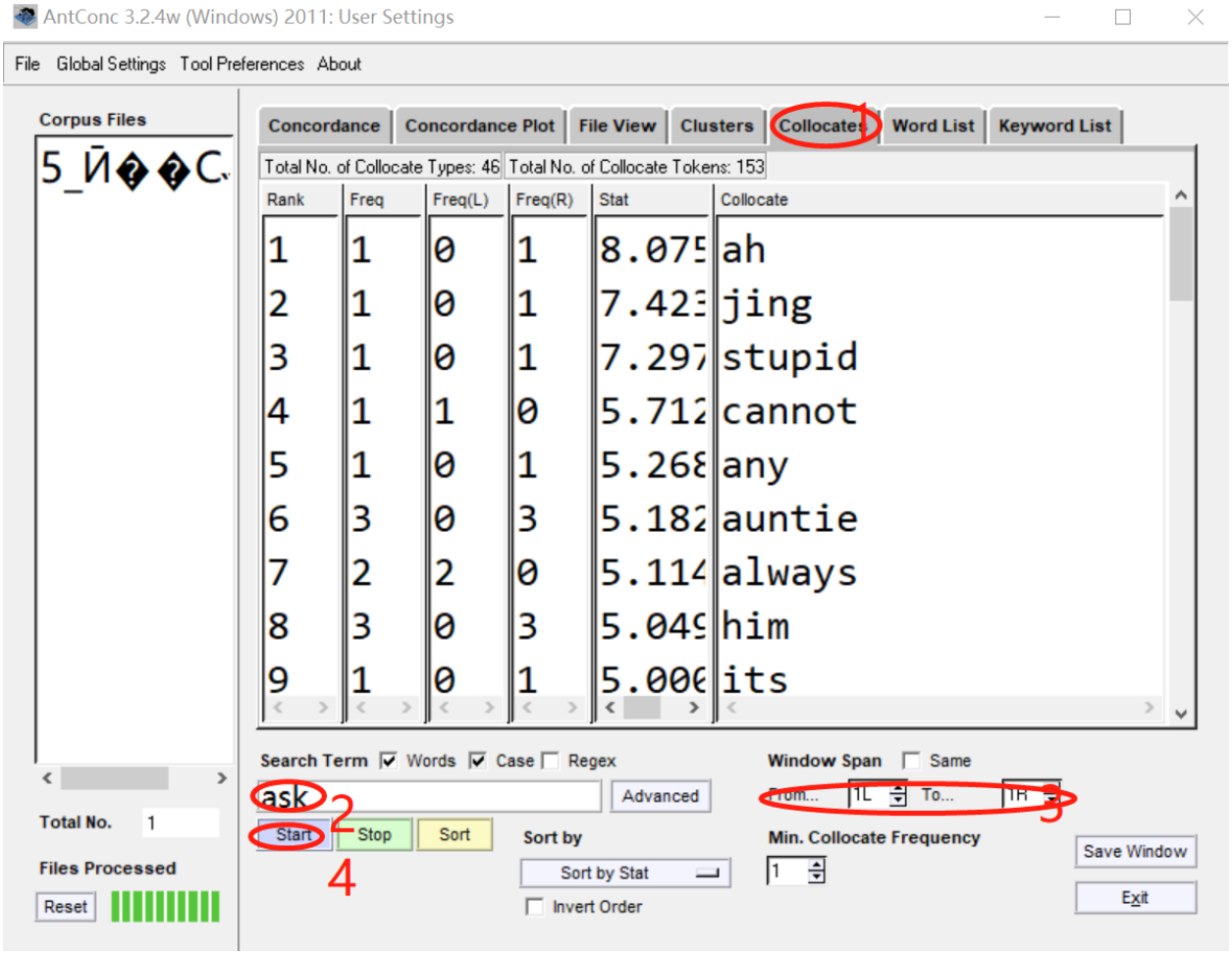
- Use the "Treat all data as lowercase" option (default) causes all words to be transformed to lowercase words. This is useful to get accurate counts of words in certain cases.
- Use the "Treat case in sort" option causes capitalized words to appear before lower-case words.

2.5 Collocates

This tool allows you to search for collocates of a search term.The collocates can be ordered either by total 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 following steps produce a set of collocates from a corpus and demonstrate the main features of this tool.

1. Click on 'Collocates'.
2. Enter a search term, like'ask'.
3. Select a span for the collocation.
4. Click on 'Start' for a search.



In addition to the above, the following settings can be made:

- Use the "Selected Collocate Measure" to choose the statistical measure for measuring collocate strength. Currently, two statistical measures can be used: Mutual Information (MI) and T-Score. See the tool explanation above for references to the statistics.
- Use the "Treat all data as lowercase" option causes all words to be transformed to lower-case words. This is useful to get accurate counts of words in certain cases.
- Use the "Treat case in sort" option causes capitalized words to appear before lower-case words.

2.6 Word List

This tool counts all the words in the corpus and presents them in an ordered list. This allows you to quickly find which words are the most frequent in a corpus.

The words can be ordered either by frequency or the start or end of the word, and the ordering 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.

The following steps produce a word list and demonstrate the main features of this tool.

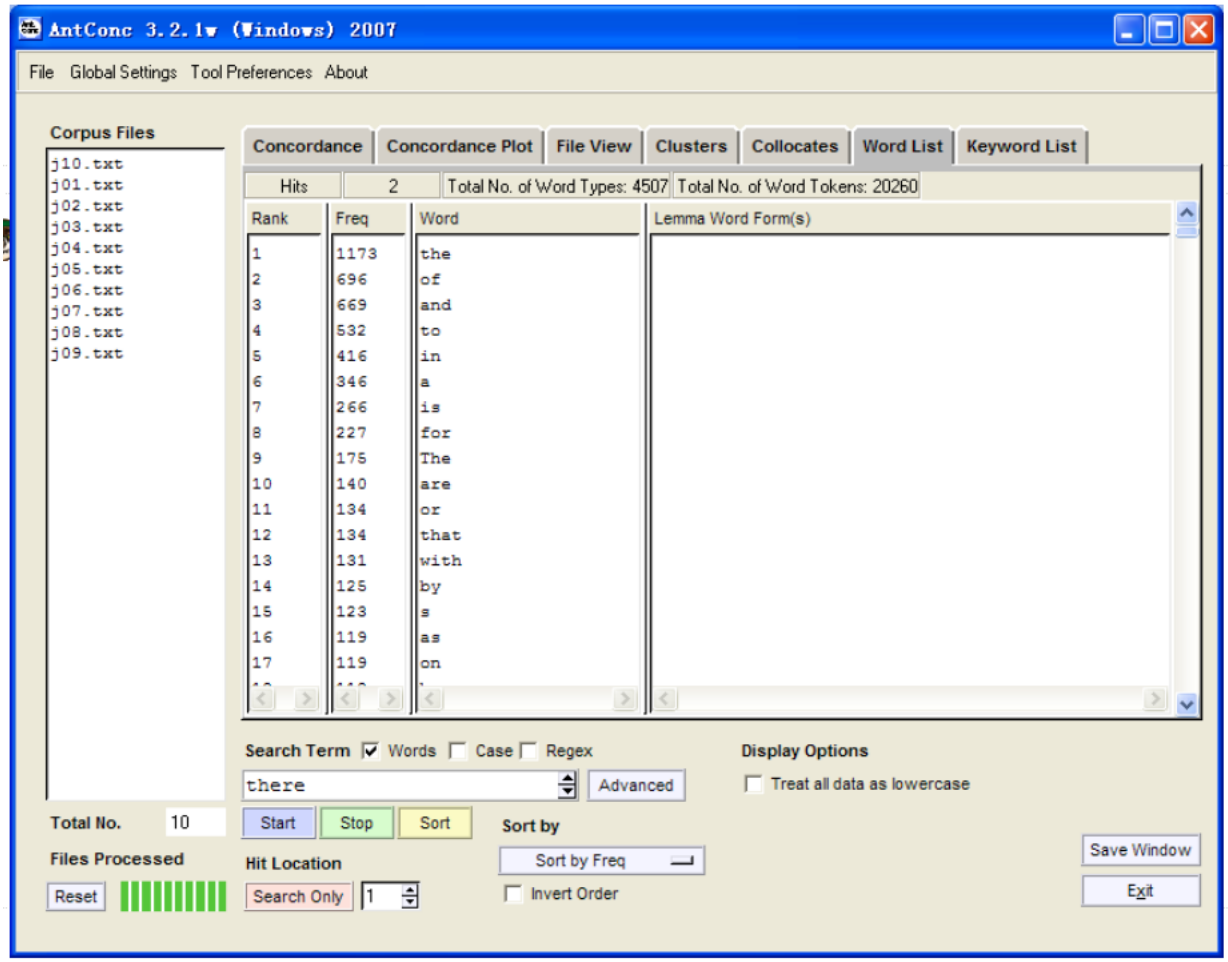
1. Choose the appropriate ordering options.

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

3. Click on the word to generate a set of KWIC lines using the text as the search term.

4. Click on the "Clone Results" button to create a copy of the results so that different sets of results can be compared.

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



2.6.1 Word List Preferences

In addition to the above the following settings can be made:

1. Using the “Treat all data as lowercase” option causes all words to be transformed to lower-case words. This is useful to get accurate counts of words in certain cases.
2. Using the “Treat case in sort” option causes capitalized words to appear before lower-case words.
3. Use the lemma list options to select a lemma list. A ‘lemma 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. A lemma list can be created by specifying the ‘lemma entry’ follow by ‘->’ followed by one or more ‘words’ that should be assigned to the lemma separated by commas or tabs (with or without additional space around the words). See the example below:

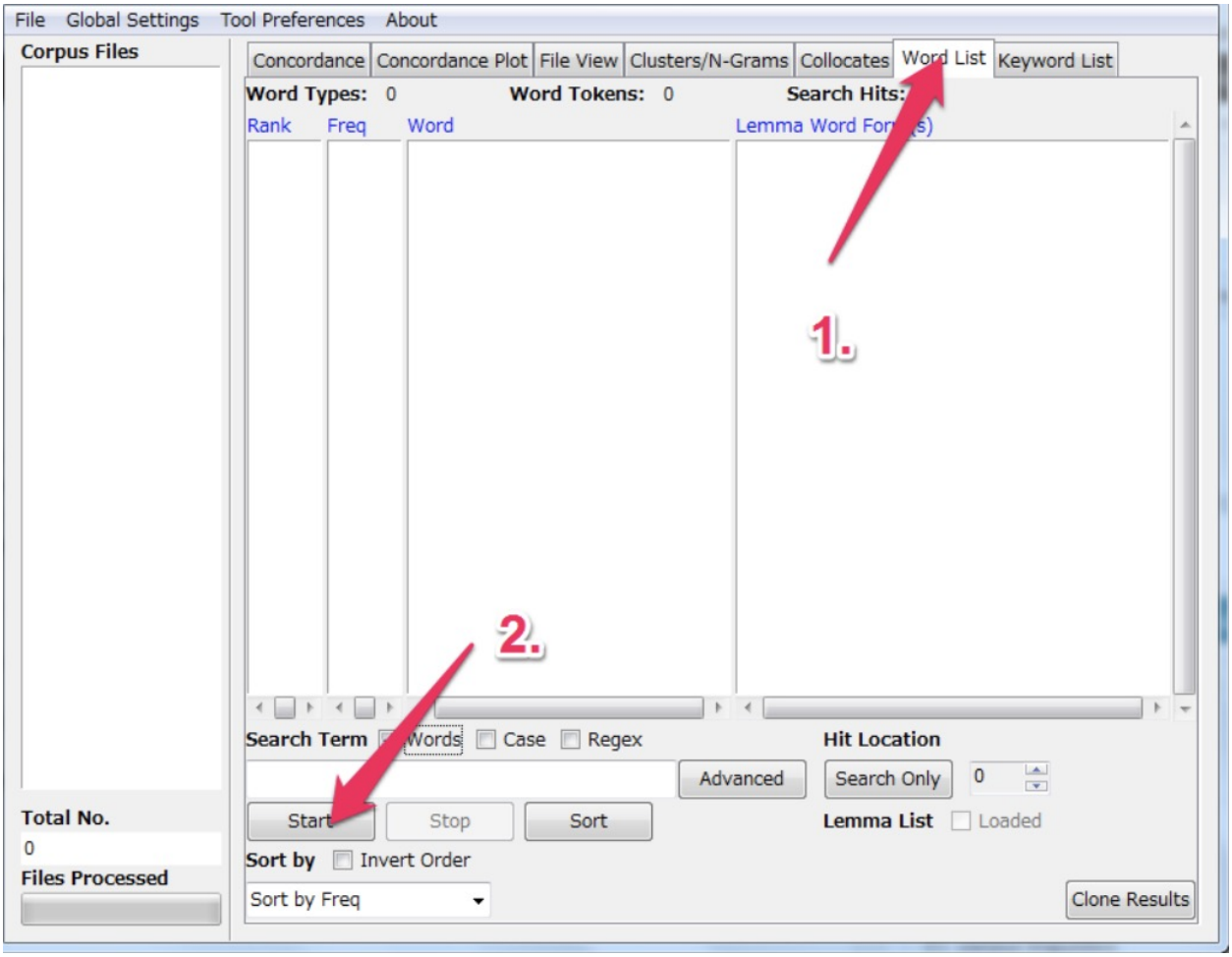
”be->is, are	be->is<tab>are”
”play -> play, plays, playing, played	play->play <tab> plays <tab> playing <tab> played”

4. Using the “Use word list” option under the “Target Corpus” heading, a pre-built word list contained in a single file can be loaded after selecting the file by clicking on “Load Word List”.
5. Using the “Word List Range” option, a wordlist can be generated using all words (“Use all words”), or a specific set of words (“Use specific words below”), or ignoring a certain set of words (“Use a stoplist below”). The range of words to be used (or ignored) can be entered directly, or can be stored in files which are then read by AntConc by pressing the ‘Open’ button. A combination of words in a file and words directly entered can also be used.

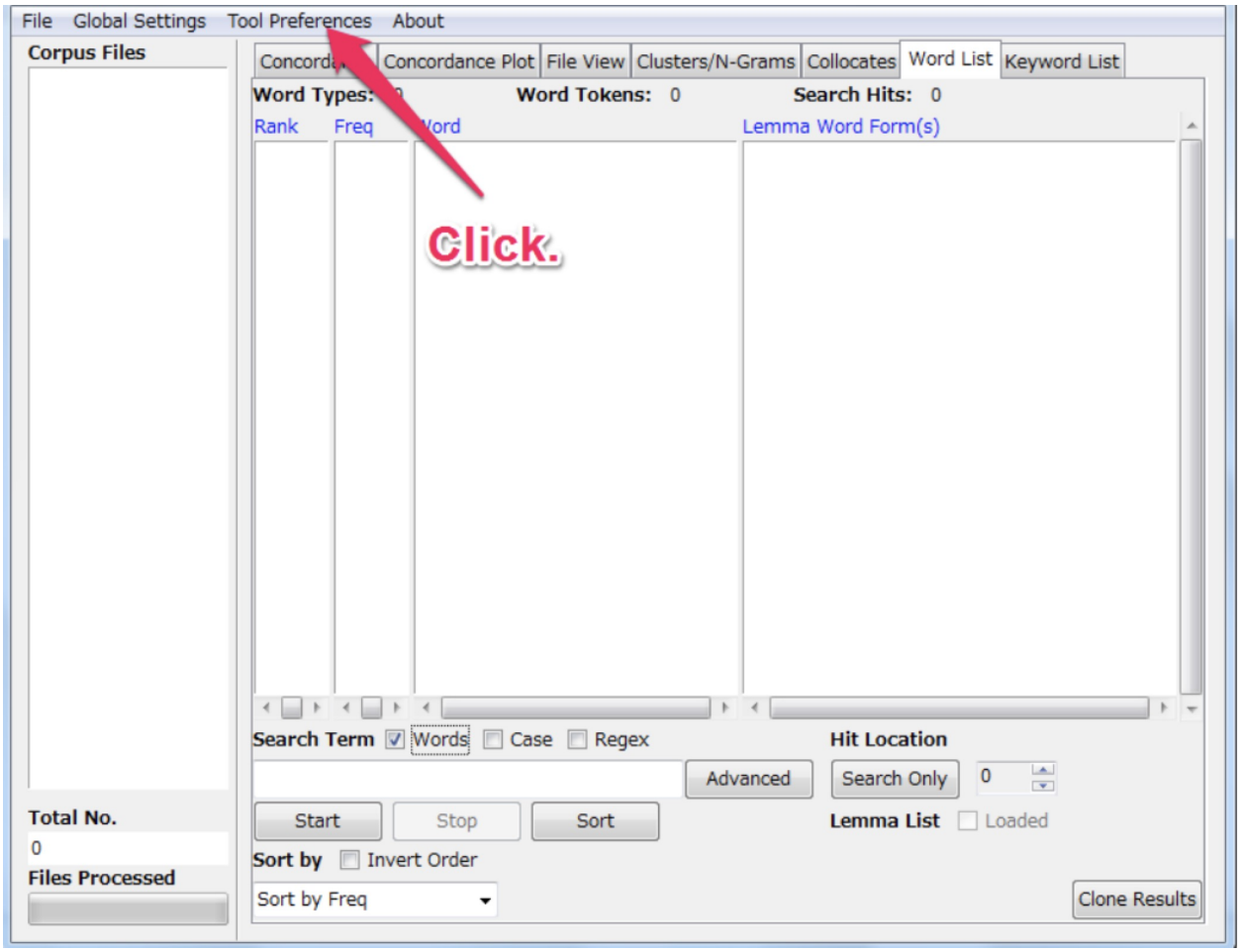
2.7 Keyword List

This tool shows the which words are unusually frequent (or infrequent) in the corpus in comparison with the words in a reference corpus. This allows you to identify characteristic words in the corpus, for example, as part of a genre or ESP study. The following steps produce a keyword list and demonstrate the main features of this tool.

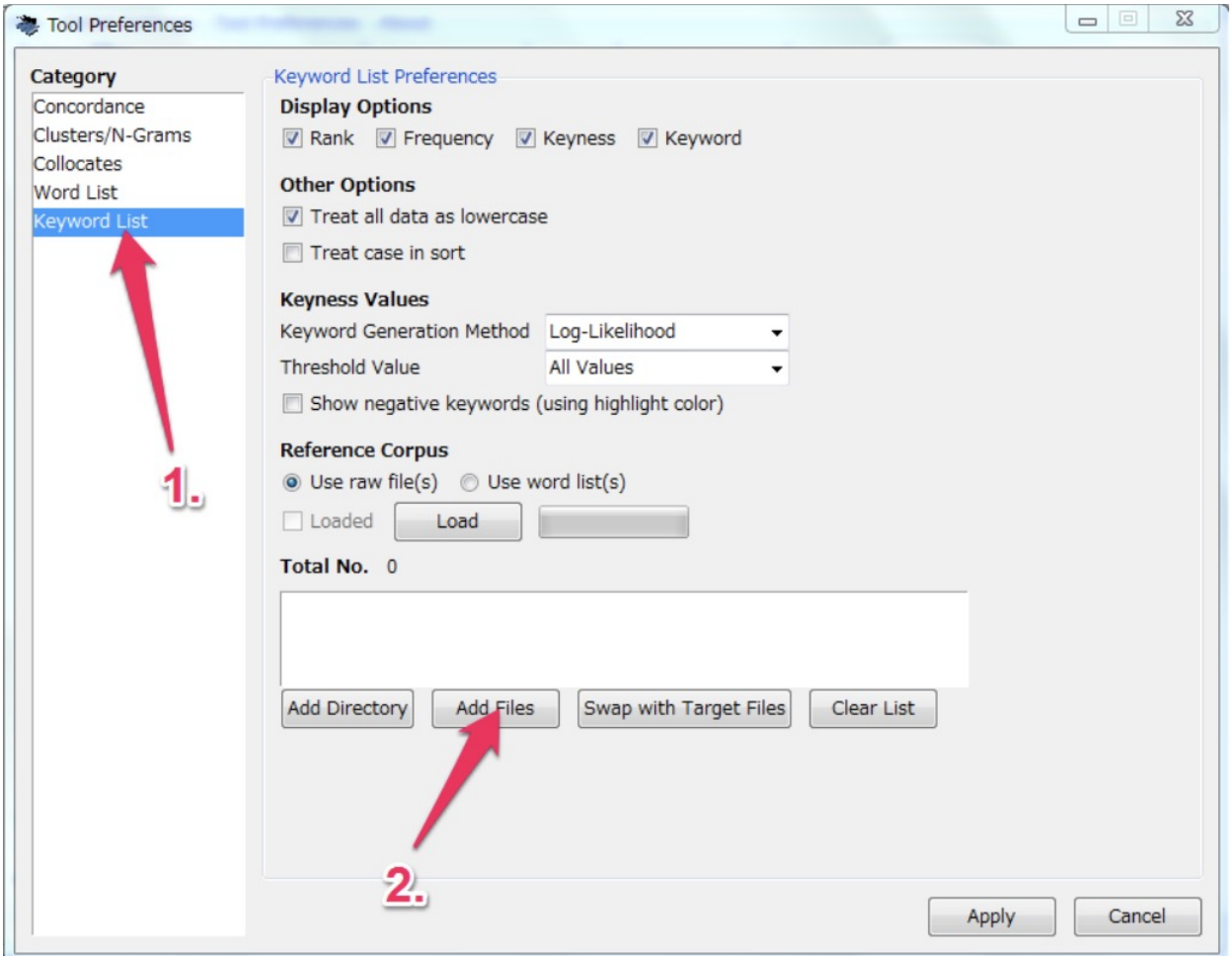
1. In AntConc load your corpus or corpora. Go to Wordlist tab then click start.



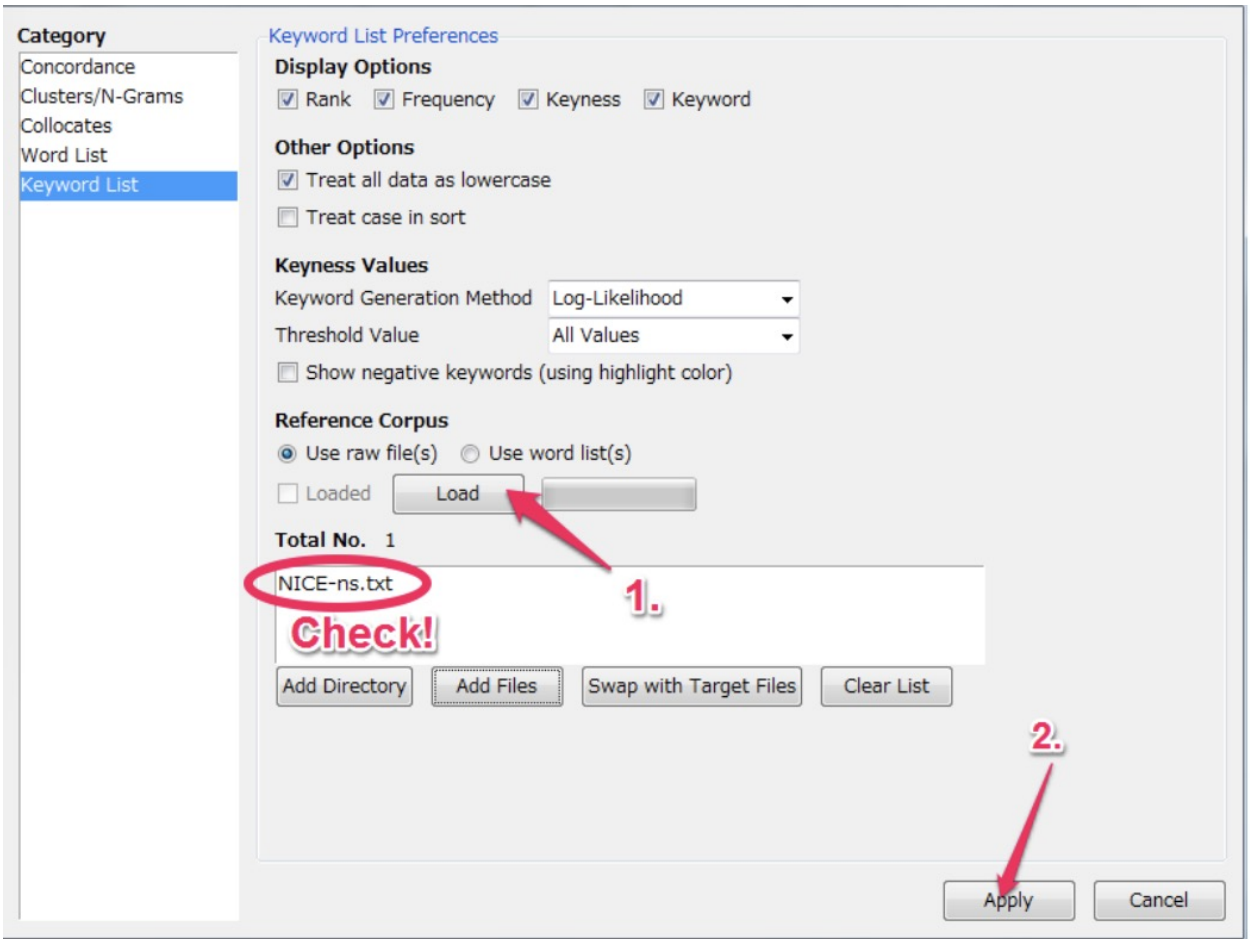
2. Select the Tools Preference menu.



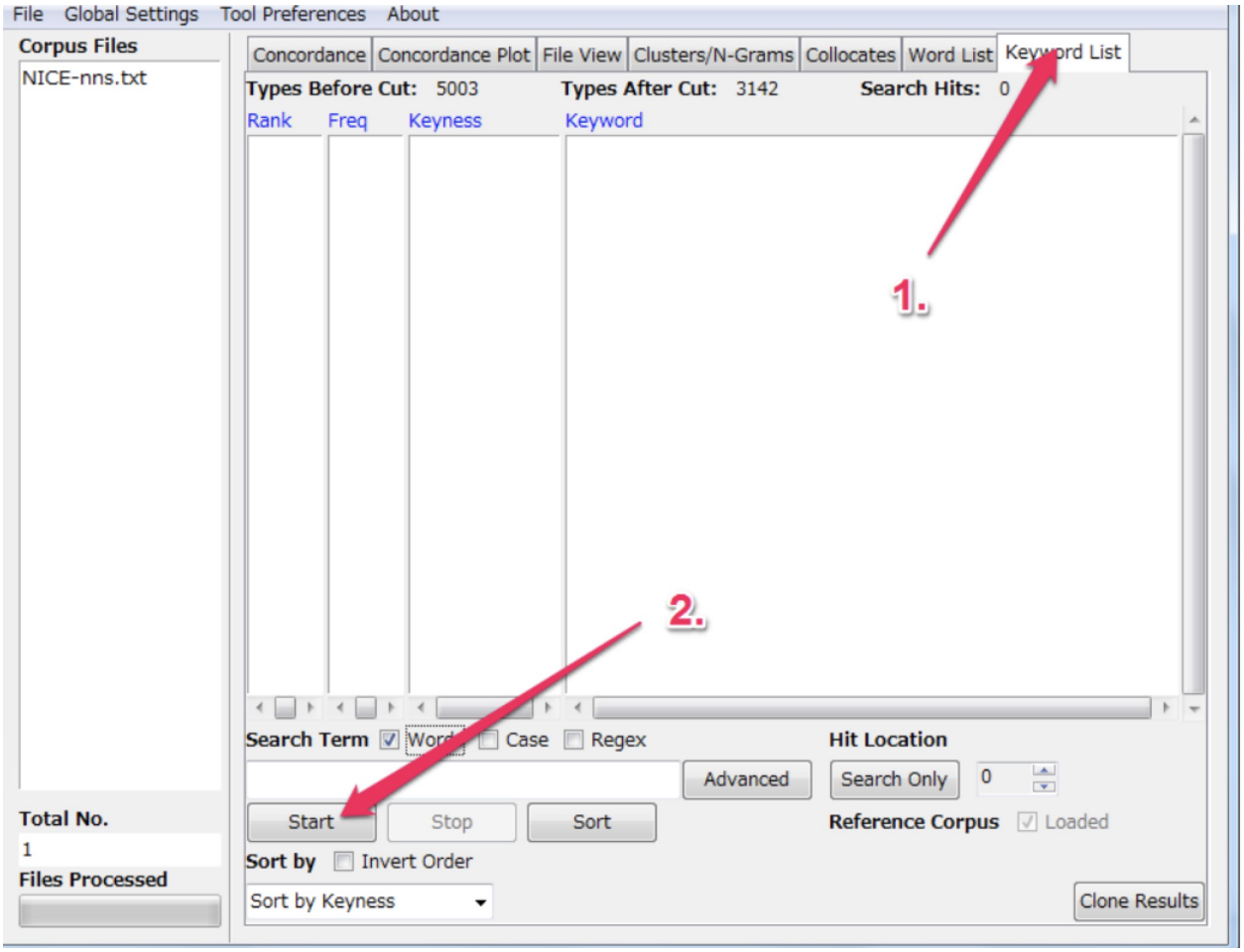
3. Click the 'Keywords List' option, then click 'Add Files'.



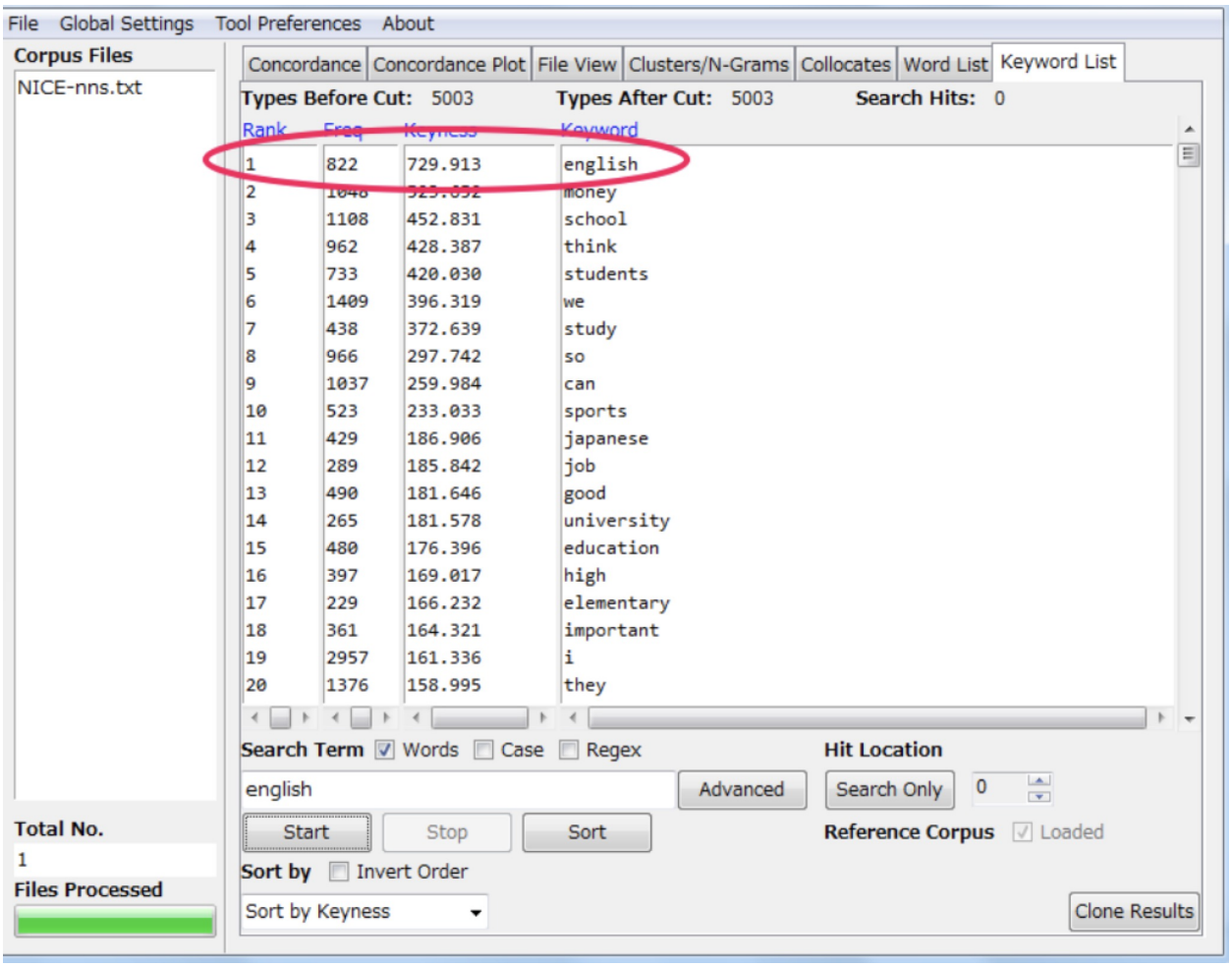
4. Check the desired file is there. Click ‘Load’ then click ‘Apply’.



5. Go to 'Keyword List' tab then click 'Start'.



6. A list of types should appear like this.



The keywords are ranked by default by the keyness. In this example the top ranking type in "english" with a score keyness (in this example, chi-squared) of 729.913 (this is a combined score of both the target and reference type score). And it has a frequency of 822 in the target list.

2.7.1 Keyword List Preferences

In addition to the above the following settings can be made:

1. Using the "Treat all data as lowercase" option causes all words to be transformed to lower-case words. This is useful to get accurate counts of words in certain cases.
2. Using the "Treat case in sort" option causes capitalized words to appear before lower-case words.
3. Use the "Show Negative Keywords" option to view words that are unusually INFREQUENT in the target corpus compared with the reference corpus.
4. Use the "Keyness Values" options to choose the statistical measure for measuring keyword strength, the statistical threshold used to decide a cutoff point, the effect size measure used to rank the keywords, and the effect size threshold to decide the cutoff of the effect size. Currently, two statistical measures can be used: Chi-Squared and Log-Likelihood (the default). The default option for the 'keyness' measure is recommended, however, you may want to experiment with different effect size measure depending on the corpus you are using. The selection of statistical measures and effect size measures are inspired by the work of Andrew Hardie of Lancaster University.
5. Use the "Use raw file(s)" option to use raw reference corpus file(s) as the reference corpus.

6. Use the "Use word list(s)" option to use word list(s) that correspond to a reference corpus. The word list(s) should be formatted as described in the tool explanation.
7. Click the "Add Directory" or "Add Files" buttons to select the reference corpus files.
8. Click the "Swap with Target Files" button to swap the main and reference corpora. Note that this will only make sense when raw corpus files are being used.

CHAPTER 3

FAQ

3.1 Facts about "Antconc"

Analysis orientation	General analysis +
Analysis type	Data mining methods and algorithms +
Data import	HTML +, TXT + and XML +
Data manipulation type	Data analysis +
End user type	Students/Learners/Consumers +, Teachers/Tutors/Managers + and Researchers +
Has completion level	Low +
Has description	AntConc is a freeware concordance program … example, as part of a genre or ESP study. +
Has interface usability	easy to use +
Has logo	**AntConc** +
Has name	AntConc +
Has screenshot	AntConc 3.4.3w(Windows)2014.jpg +
Has website	http://www.laurenceanthony.net/software.html#antconc +
License type	Free&Closed source +
Mining tool type	Text mining +
User data mining models level	N/A +
User programming level	N/A +
User statistics level	N/A +
User system engineering level	N/A +

3.2 About Users

Tool is suitable for: Students/Learners/Consumers Teachers/Tutors/Managers Researchers

Tool is unsuitable for: Developers/Designers Organisations/Institutions/Firms Others

3.3 Performance Issues

The performance of searches can be very slow when the case option is deactivated (the default). Much higher performance can be gained by activating this option. (The performance drop here is related to the handling of Unicode characters). One further way to improve performance (when doing multi-word searches only) is to deactivate the "Treat search whitespace as one or more non-tokens" in the Wildcard global settings. (Again, the performance drop here is related to the handling of Unicode characters).

3.4 About Shortcuts

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

1. CTRL-C (COMMAND-C on Macintosh OS X) = Copies the currently selected text
2. CTRL-A (COMMAND-A on Macintosh OS X) = Selects all text in the window pane
3. ALT-A (OPTION-A or COMMAND+OPTION-A on Macintosh OS X) = Selects all text in all window panes showing
4. Double click = Selects the current word
5. Triple click = Selects the current line in the window pane
6. SHIFT-click = Selects continuous lines across all window panes showing CTRL-click = Selects discontinuous lines across all window panes showing
7. DELETE (BACKSPACE on Macintosh OS X) = This deletes any selected lines that span across all window panes INSERT (SHIFT + BACKSPACE on Macintosh OS X) = This keeps any selected lines that span across all window panes, and deletes all others
8. For any 'spinbox' widgets (e.g. the search term entry box) the 'UP' and 'DOWN' arrow keys on the keyboard can be used to activate the up and down buttons.

3.5 Saving Results

Results can be saved to the clipboard, saved to a text file (.txt), saved to a postscript file (.ps) -for the Concordance Plot tool, or saved to a new window using keyboard commands, the appropriate option in the 'File Menu', or by clicking on the "Save Window" button in each tool, respectively. Also, it is possible to launch multiple clones of AntConc by double clicking on the .exe file.

3.6 Comments/Suggestions/Bug Fixes

All new editions and bug fixes are listed in the revision history below. However, if you find 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 a 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.

CHAPTER 4

References

Below are some links you might find useful.

- Official websiate laurenceanthony.net
- What AntConc's Concordance Tool Can Do For You HECC
- A Guide to using AntConc research.ncl.ac.uk
- AntConc Walk-Through The Grammar Lab
- YouTube Tutorial YouTube Tutorial
- Edutechwiki Edutechwiki