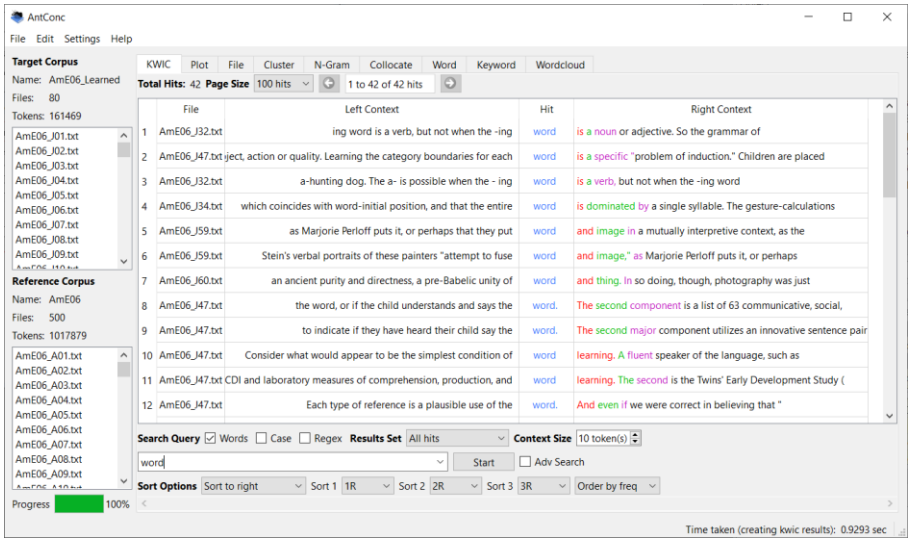
AntConc (Windows, MacOS, Linux)

Build 4.1.2

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
September 25, 2022



Introduction

AntConc is a freeware, multiplatform tool for carrying out corpus linguistics research, introducing corpus methods, and doing data-driven language learning. It runs on any computer running Microsoft Windows (built on Win 10), MacOS (built on Mac Catalina), and Linux (built on Linux Mint). It is developed in *Python* and *Qt* using the *PyInstaller* compiler to generate executables for the different operating systems. It uses SQLite as the underlying database.

Getting Started

Windows - Installer

Double click the *AntConc.exe* file and follow the instructions to install the application into your Programs folder. You can delete the .exe file when you are finished. You can start the application via the Start Menu.

Windows - Portable

Unzip the *AntConc.zip* file into a folder of your choice. In the *AntConc* folder, double click the *AntConc.exe* file to launch the program.

Macintosh OS X

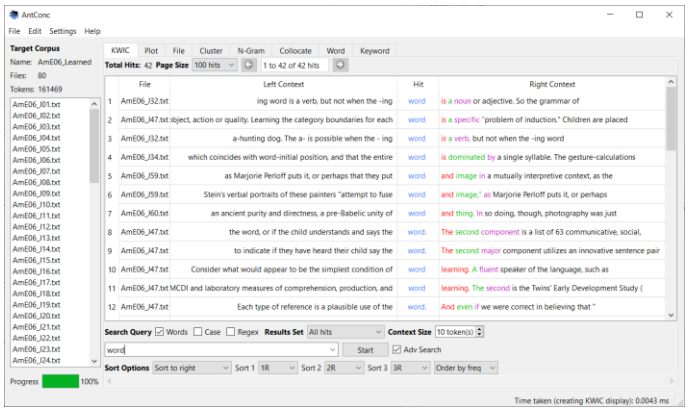
Double click the *AntConc.dmg* file to create a AntConc disk image on your desktop. Open the disk image and drag and drop the AntConc app onto the Applications folder (or into another location if you desire). You can then launch the app by double clicking on the icon in the Applications folder or the Launchpad.

Linux

Decompress the *AntConc.tar.gz* file into a folder of your choice. In the *AntConc* folder, double click the *AntConc.sh* file to launch the software. On the command line, type ./AntConc.sh to launch the software.

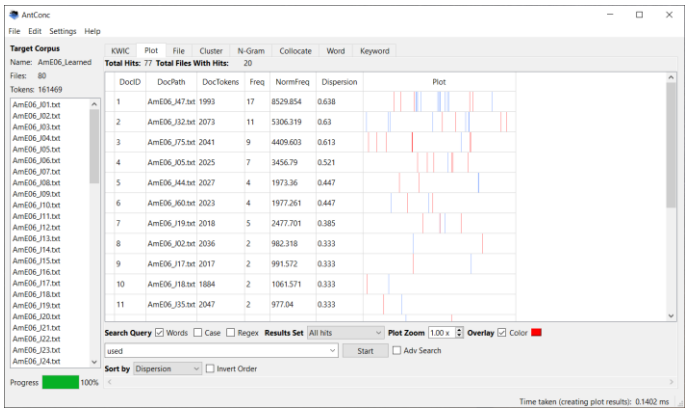
Overview of Tools

AntConc contains nine tools that can be accessed either by clicking on their 'tabs' in the tool window, using CTRL+TAB to toggle through the tools, or using the key combination CTRL + Tool Number (e.g., CTRL +1 for KWIC, CTRL +2 for Plot) to select a specific tool.



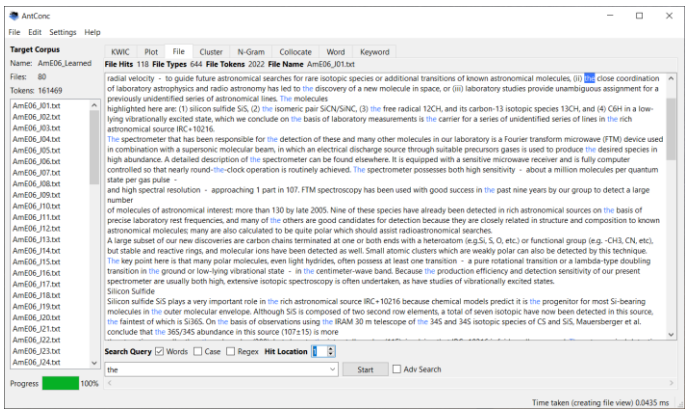
KWIC (Key-Word-In-Context) Tool

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



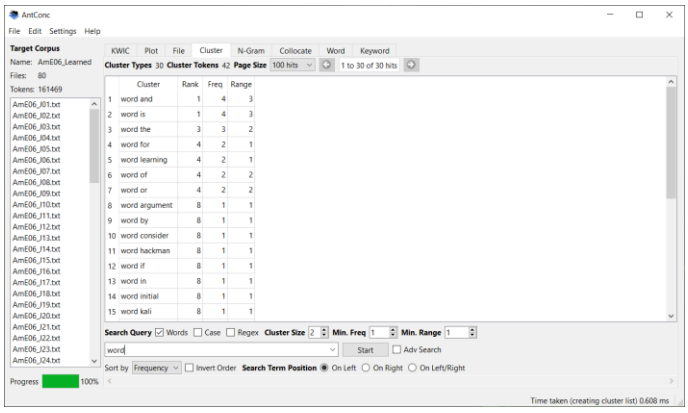
Plot Tool

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 the individual texts of a corpus.



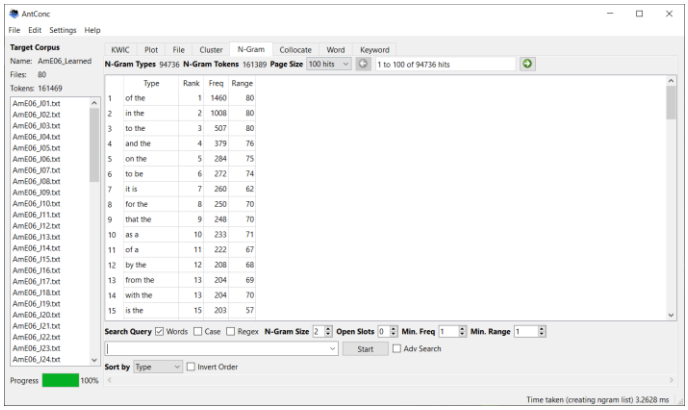
File Tool

This tool shows the contents of individual texts. This allows you to investigate in more detail the results generated in other tools of *AntConc*.



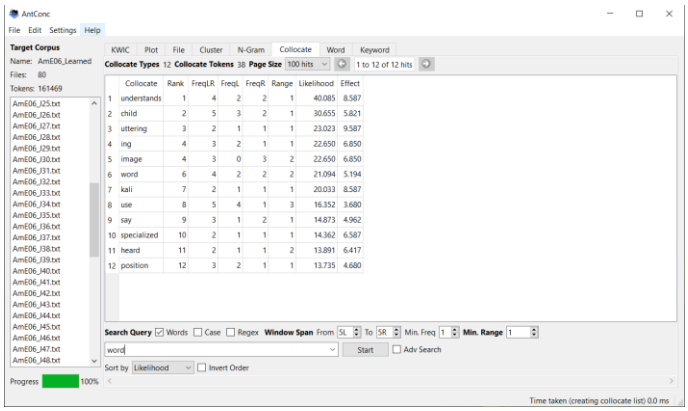
Cluster Tool

The tool shows contiguous (together in a sequence) word patterns based on the search condition. This allows you to see common phrases that appear in the target texts.



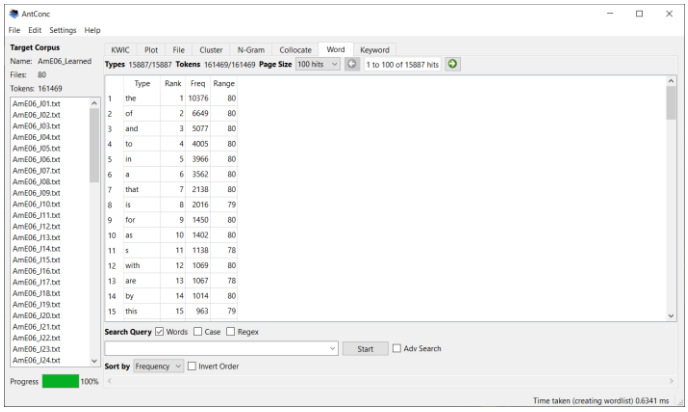
N-Gram Tool

This tool scans the entire corpus for all 'N'-sized clusters (e.g., 2-word clusters, 3-word clusters, …). This allows you to find common expressions in a corpus.



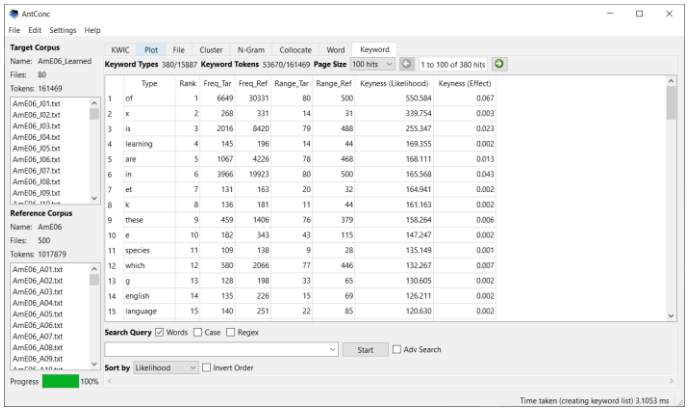
Collocate Tool

This tool shows words that appear frequently within a certain distance of the search term (i.e., collocates). This allows you to find which words co-occur with other words in a corpus.



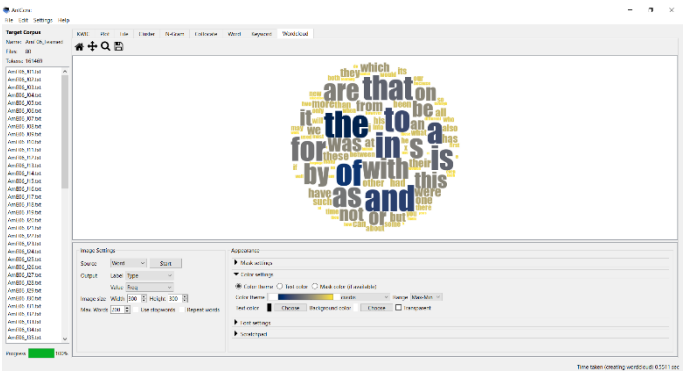
Word List Tool:

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



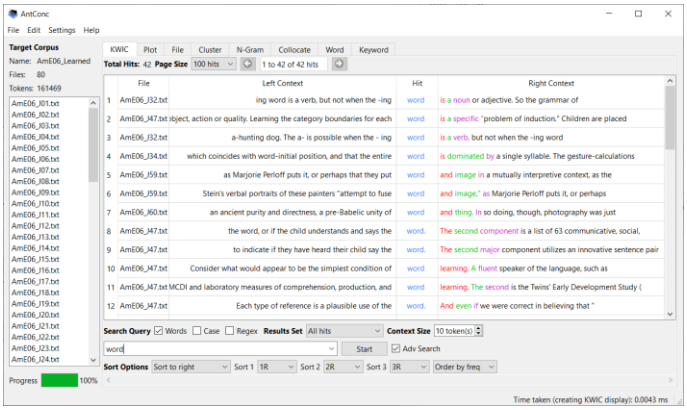
Keyword List Tool:

This tool shows words that appear unusually frequently in the target corpus in comparison with the words in the reference corpus based on a statistical measure (i.e., 'keywords'). These words can be considered to be characteristic of the target corpus. The settings can also be changed to show words that appears unusually infrequently in the target corpus compared with the reference corpus (i.e., 'negative keywords'.



Wordcloud Tool:
This tool visualizes the results generated by KWIC, File, Cluster, N-Gram, Collocate, Word, and Keyword tools as well as a “Scratchpad” of plain text in the form of a ‘word cloud’. Wordclouds are often used as aesthetically pleasing visualizations, where words are laid out in a viewing area or ‘themed’ image mask and sized according to a property (e.g., word frequency). Care should be taken when using wordclouds for linguistic analysis, as the visualization necessitates distorting word sizes to fit the viewing area.

KWIC (' Key-Word-In-Context') Tool



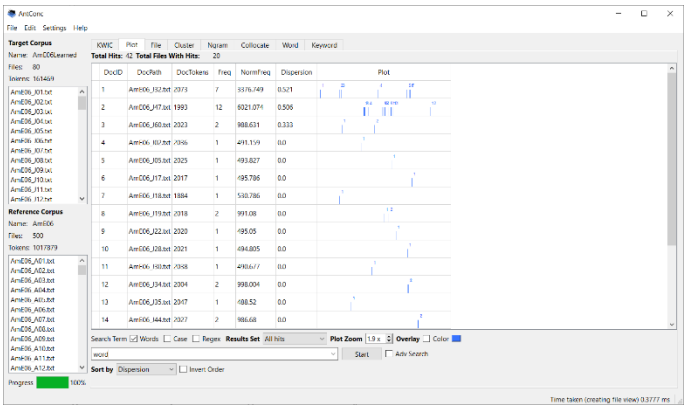
This tool shows search results in a concordance or 'KWIC' (Key-Word-In-Context) format. This allows you to see how words and phrases are commonly used in a corpus of texts.

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

1) Select a corpus using the "Corpus Manager" available from the File menu. Alternatively, create a quick corpus by choosing the "Open File(s) as Quick Corpus" from the file menu. The files contained in the corpus are shown in the left frame of the main window under "Target Corpus".
2) Enter a search query in the search box. See the 'SEARCH OPTIONS' section in this document for an explanation of the "Words", "Case", and "Regex" search term options.
3) Choose the size of the results set to be presented using the "Result Set" combobox widget.
4) Choose the number of words to be displayed on either side of the search term using the "Context Size" spinbox widget.
5) Click on the "Start" button to start the search and wait for the results to be displayed.
6) Use the "Sort options" to rearrange the concordance lines by row ID, file name, or the position of the word. The first widget allows you to quickly order the concordance lines by the words to right or left of the center word, or choose no ordering, or using a custom order. The next three widgets allow you to choose the order parameters: 1L, 2L... are words to the left of the target word, 'C' is the center word, and 1R, 2R... are words to the right of the center word. The final widget allows you to order the results by the frequency of the pattern determined by the sorting parameters (the "Order by freq" option) or alphabetically (the "Order by value" option). The default "Order by freq" option is strongly recommended as it will allow you to easily identify the most commonly occurring patterns in the target corpus. After adjusting the sort options, click on the "Start" button to regenerate the concordance lines.
7) The total number of concordance lines generated (Total Hits) is shown at the top of the tool window. When no hits are found, a warning will be shown on the screen.
8) Double-clicking on any cell in the results window will cause the software to jump to the File tool (see the relevant section of this document) where you can view the hit exactly as it appears in context in the original file.
9) If you want to filter the results, select the desired rows and then press the "Delete" key to remove the selected rows or press "SHIFT+ Delete" to keep the selected rows removing all the others.

10) Advanced searches are available with this tool. Several menu preferences are also available with this tool. (See the relevant sections in this document for explanations).

Plot Tool

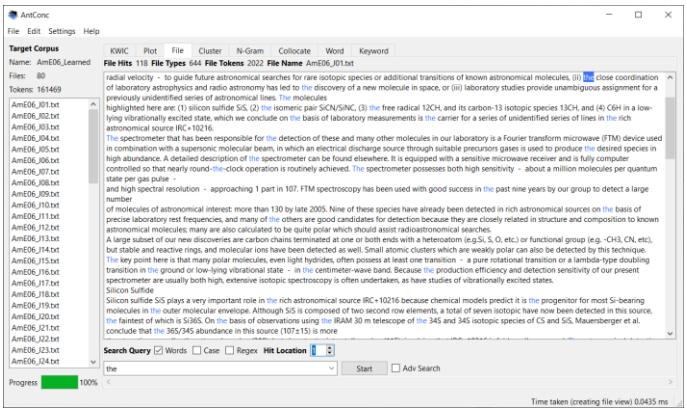


This tool shows concordance search results plotted in a 'barcode' format, with the length of the text normalized to the width of the bar and each hit shown as a vertical line within the bar. This allows you to see the position where search results appear in the individual texts of a corpus. An example of the use of the Plot Tool is in determining where specific content words appear in a technical paper, or where an actor or story character appears through a play or novel.

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

1) Select a corpus using the "Corpus Manager" available from the File menu. Alternatively, create a quick corpus by choosing the "Open File(s) as Quick Corpus" from the file menu. The files contained in the corpus are shown in the left frame of the main window under "Target Corpus".
2) Enter a search query in the search box. See the 'SEARCH OPTIONS' section in this document for an explanation of the "Words", "Case", and "Regex" search term options.
3) Choose the size of the results set to be presented using the "Result Set" combobox widget.
4) Use the "Plot Zoom" widget to control the size of the plot and the degree of detail to be shown.
5) Click on the "Start" button to start the search and wait for the results to be displayed.
6) Use the "Sort by" option to rearrange the plots according to the various parameters shown.
7) The total number of hits and total number of plots are shown at the top of the tool window. When no hits are found, a warning will be shown on the screen.
8) Double-clicking on any cell in the results window will cause the software to jump to the File tool (see the relevant section of this document) where you can view the hit exactly as it appears in context in the original file.
9) By checking the "Overlay" option and choosing an appropriate color (by clicking on the color box), existing results can be overlaid with new results for different searches. This allows you to see how different search queries are related and/or overlap.
10) Advanced searches are available with this tool. Several menu preferences are also available with this tool. (See the relevant sections in this document for explanations).

File Tool



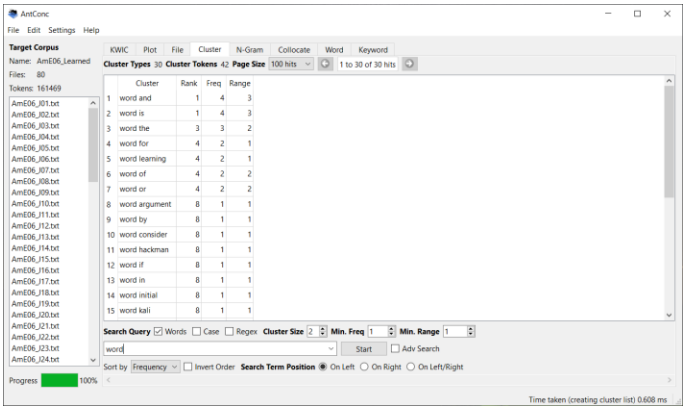
This tool shows the text of individual files. This allows you to investigate in more detail the results generated in other tools of *AntConc*.

The following steps produce a view of the original file and demonstrate the main features of this tool.

1) Select a corpus using the "Corpus Manager" available from the File menu. Alternatively, create a quick corpus by choosing the "Open File(s) as Quick Corpus" from the file menu. The files contained in the corpus are shown in the left frame of the main window under "Target Corpus".

2) Double-click a file in the “Target Corpus” list on the left of the main window to view its contents. Alternatively, select a file in the "Target Corpus" list and click "Start" in the tool interface. The File tool will automatically be selected, and the contents of the file will be shown.
3) To highlight particular search query results in the display, enter a search query and click "Start". See the relevant section in this document explaining the "Words", "Case", and "Regex" search term options. Words in the file that match the query will be automatically highlighted.
4) Use the "Hit Location" widget to jump to different hits in the file. Alternatively, use the keyboard shortcut for your operating system (see the SHORTCUTS section).
5) Advanced searches are available with this tool. Several menu preferences are also available with this tool. (See the relevant sections in this document for explanations).

Cluster Tool

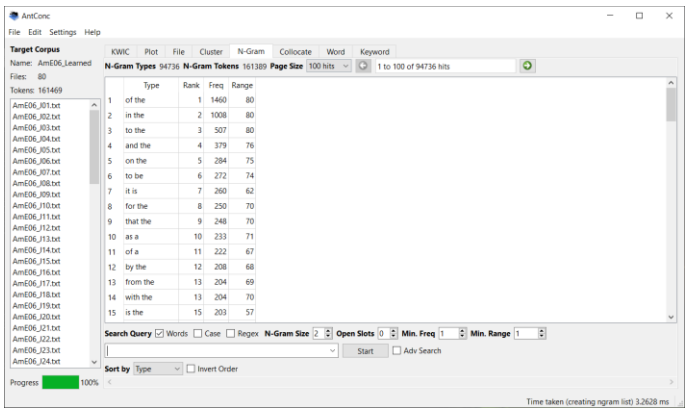


The tool shows adjacent word groups based on the search condition. This allows you to see how words and phrases are commonly used in a corpus of texts. In some cases, this tool can be seen as summarizing the results generated in the KWIC tool

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

1) Select a corpus using the "Corpus Manager" available from the File menu. Alternatively, create a quick corpus by choosing the "Open File(s) as Quick Corpus" from the file menu. The files contained in the corpus are shown in the left frame of the main window under "Target Corpus".
2) Enter a search query in the search box. See the 'SEARCH OPTIONS' section in this document for an explanation of the "Words", "Case", and "Regex" search term options.
3) Choose the various parameters to filter the number of clusters to be shown: cluster size (number of words in the cluster), minimum cluster frequency, and minimum cluster range (number of files)
4) Click the "Start" button to start the search and wait for the results to be displayed.
5) Use the "Sort by" option to rearrange the ordering of the results.
6) Use the "Search Term Position" options to determine if the results will show clusters that start with the search query terms ("On Left"), end with the search query terms ("On Right") or can either start or end with the search query terms ("On Left/Right").
7) The total number of cluster types ("Cluster Types") and combined total count of all the cluster tokens ("Cluster Tokens") are shown at the top of the tool window. When no hits are found, a warning will be shown on the screen.
8) Double-click on any cell in the results window to cause the software to jump to the KWIC tool (see the relevant section of this document) where you can view concordance lines for that cluster across the whole corpus.
9) Advanced searches are available with this tool. Several menu preferences are also available with this tool. (See the relevant sections in this document for explanations).

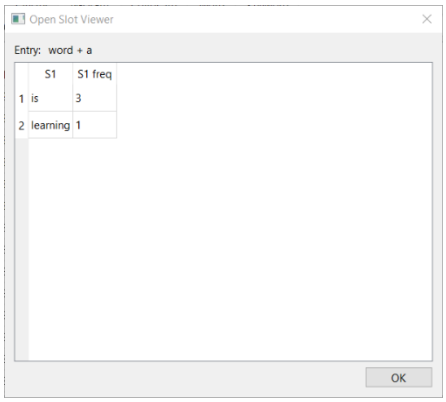
N-Gram Tool



This tool scans the entire corpus for all 'N'-sized clusters (e.g., 2-word clusters, 3-word clusters, …). This allows you to find common expressions in a corpus.

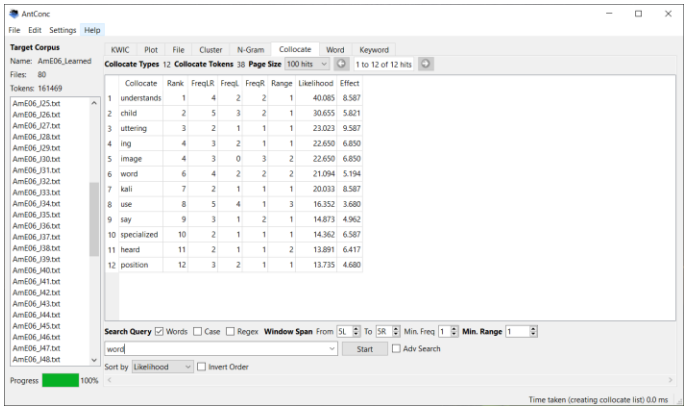
The following steps produce a set of n-grams and demonstrate the main features of this tool.

1) Select a corpus using the "Corpus Manager" available from the File menu. Alternatively, create a quick corpus by choosing the "Open File(s) as Quick Corpus" from the file menu. The files contained in the corpus are shown in the left frame of the main window under "Target Corpus".
2) Choose the various parameters to filter the number of n-grams to be shown: n-gram size (number of words), open slots (number of slots in the n-gram that can take multiple values), minimum n-gram frequency, and minimum n-gram range (number of files).
3) Click on the "Start" button to start the search and wait for the results to be displayed. If a search query is entered, only n-grams that match the query will be shown. See the 'SEARCH OPTIONS' section in this document for an explanation of the "Words", "Case", and "Regex" search term options.
4) Use the "Sort by" option to rearrange the ordering of the results.
5) The total number of n-gram types ("N-Gram types") and combined total count of all the n-gram tokens ("N-Gram Tokens") are shown at the top of the tool window. When no hits are found, a warning will be shown on the screen.
6) Double-click on any cell in the results window to cause the software to jump to the KWIC tool (see the relevant section of this document) where you can view concordance lines for that n-gram across the whole corpus.
7) For entries that contain open slots, Shift + Double-click on the "Type" entry in the results window to show the variants that can fit in the open slots via the "Open Slot Viewer" and two associated statistics that show the degree of variation for the slot. The *_TT value is the type/token ratio for the slot, and the *_ent value is the Entropy value for the slot.
8) Advanced searches are available with this tool. Several menu preferences are also available with this tool. (See the relevant sections in this document for explanations).



Collocate Tool

This tool shows words that appear frequently within a certain distance of the search term (i.e., collocates). This allows you to find which words co-occur with other words in a corpus.

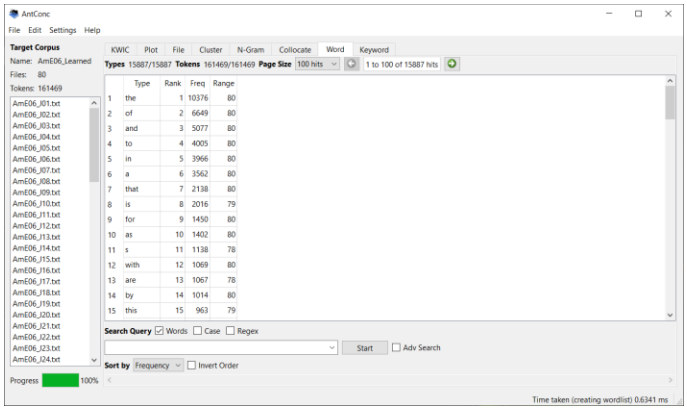


The following steps produce a set of collocates and demonstrate the main features of this tool.

1) Select a corpus using the "Corpus Manager" available from the File menu. Alternatively, create a quick corpus by choosing the "Open File(s) as Quick Corpus" from the file menu. The files contained in the corpus are shown in the left frame of the main window under "Target Corpus".
2) Choose the various parameters to filter the types of collocates to be shown: window span (possible positions left and right of the search query terms, where the collocate can appear), minimum collocate frequency, and minimum collocate range (number of files).
3) Enter a search query in the search box. See the 'SEARCH OPTIONS' section in this document for an explanation of the "Words", "Case", and "Regex" search term options.
4) Click on the "Start" button to start the search and wait for the results to be displayed.
5) Use the "Sort by" option to rearrange the ordering of the results.

6) The total number of collocate types ("Collocate Types") and combined total count of all the collocate tokens ("Collocate Tokens") are shown at the top of the tool window. When no hits are found, a warning will be shown on the screen.
7) Double-click on any cell in the results window to cause the software to jump to the KWIC tool (see the relevant section of this document) where you can view concordance lines for that collocate across the whole corpus.
8) Advanced searches are available with this tool. Several menu preferences are also available with this tool. (See the relevant sections in this document for explanations).

Word Tool

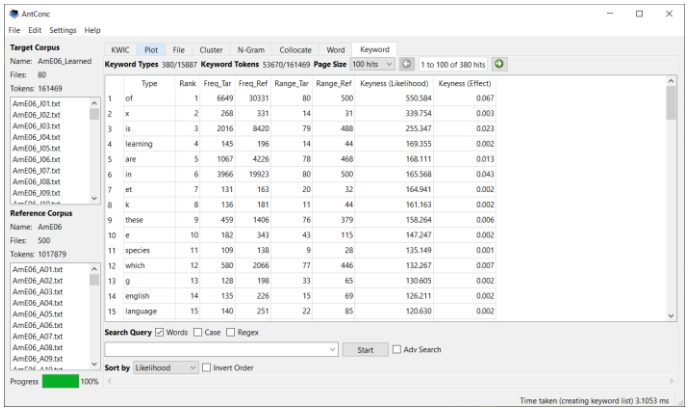


This tool counts all the words in the corpus and presents them in an ordered list. This allows you to find which words are the most frequent in a corpus.

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

1) Select a corpus using the "Corpus Manager" available from the File menu. Alternatively, create a quick corpus by choosing the "Open File(s) as Quick Corpus" from the file menu. The files contained in the corpus are shown in the left frame of the main window under "Target Corpus".
2) Click on the "Start" button to start the processing and wait for the results to be displayed. If a search query is entered, only words that match the query will be shown. See the 'SEARCH OPTIONS' section in this document for an explanation of the "Words", "Case", and "Regex" search term options.
3) Use the "Sort by" option to rearrange the ordering of the results.
4) The total number word types ("Word types") and combined total count of all the word tokens ("Word Tokens") are shown at the top of the tool window. When no hits are found, a warning will be shown on the screen.
5) Double-click on any cell in the results window to cause the software to jump to the KWIC tool (see the relevant section of this document) where you can view concordance lines for that word across the whole corpus.
6) Advanced searches are available with this tool. Several menu preferences are also available with this tool. (See the relevant sections in this document for explanations).

Keyword Tool



This tool shows words that appear unusually frequently in the target corpus in comparison with the words in the reference corpus based on a statistical measure (i.e., 'keywords'). These words can be considered to be characteristic of the target corpus. The settings can also be changed to show words that appears unusually infrequently in the target corpus compared with the reference corpus (i.e., 'negative keywords'.

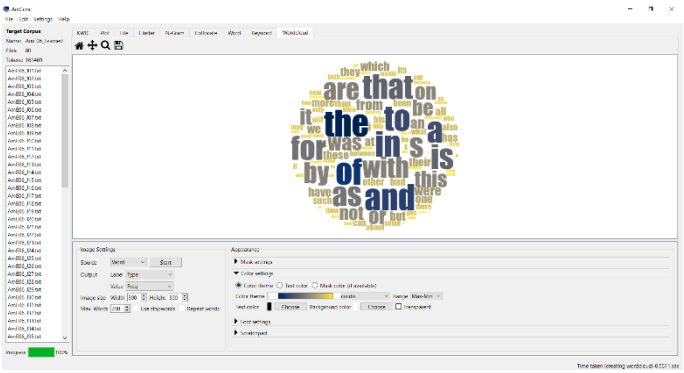
The following steps produce a keyword list and demonstrate the main features of this tool.

1) Create a quick corpus by choosing the "Open File(s) as Quick Corpus" option from the file menu. Alternatively, choose the "Corpus Manager" option from the file menu and make sure the "Target

Corpus" option is selected. Then, select one of the available corpora or create your own from raw files or a word list (see the instruction under the Corpus Manager section of this help page for how to do this). This corpus will then serve as the target corpus for your analysis. The files contained in the corpus will be shown in the top left frame of the main window under "Target Corpus".

2) Choose a reference corpus by opening the "Corpus Manager" option from the file menu and checking the "Reference Option" option. Next, as in step 1, select one of the available corpora or create your own from raw files or a word list. The files contained in the corpus will be shown in the bottom left frame of the main window under "Reference Corpus".
3) Click on the "Start" button to start the processing and wait for the results to be displayed. If a search query is entered, only words that match the query will be shown. See the 'SEARCH OPTIONS' section in this document for an explanation of the "Words", "Case", and "Regex" search term options.
4) Use the "Sort by" option to rearrange the ordering of the results.
5) The total number keyword types ("Keyword types") and combined total count of all the word tokens ("Keyword Tokens") are shown at the top of the tool window. When no hits are found, a warning will be shown on the screen.
6) Double-click on any cell in the results window to cause the software to jump to the KWIC tool (see the relevant section of this document) where you can view concordance lines for that keyword across the whole corpus.
7) Advanced searches are available with this tool. Several menu preferences are also available with this tool. (See the relevant sections in this document for explanations).

Wordcloud Tool



This tool visualizes the results generated by KWIC, File, Cluster, N-Gram, Collocate, Word, and Keyword tools as well as a “Scratchpad” of plain text in the form of a ‘word cloud’. This is a grouping of words where the sizing of the words reflects a property of those words (e.g., frequency). Wordclouds are often used as aesthetically pleasing visualizations, where words are laid out in a viewing area or ‘themed’ image mask and sized according to a property (e.g., word frequency). Care should be taken when using wordclouds for linguistic analysis, as the visualization necessitates distorting word sizes to fit the viewing area.

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

1. Choose a “Source” for the word cloud. This can be the output of the KWIC, File, Cluster, N-Gram, Collocate, Word, and Keyword tools, or a “Scratchpad” of plain text (accessible in the bottom right of righthand windowpane).
2. Choose the properties from the source to display as “labels” and “values” in the word cloud. For the KWIC, File and Scratchpad sources, the properties of “Type” (Word) and “Freq” (Frequency) are chosen automatically.
3. Choose the image size. This will determine how many words can be placed in the image. Note that the image will be automatically scaled to the window display size.
4. Choose the maximum number of words to display in the word cloud. Depending on the number of items from the source and the other settings (e.g., the minimum font size), this value might not be reached.
5. Check the “Use stopwords” checkbox to remove stopwords from the word cloud.
6. Check the “Repeat words” checkbox to fill a much remaining space as possible in the word cloud image with existing words. Depending on other settings (e.g., the minimum font size), not all the space will be used.

7. Click “Start” to generate the word cloud.

Appearance Option

The appearance of the word cloud can be adjusted through the following settings:

Mask settings: Use these settings to determine if the word cloud should be ‘masked’ and if so, which mask to use. Additional masks (.png and .svg files) can be added to the list by clicking the “Add” button. The included masks with .svg extensions are kindly provided by Font Awesome (fontawesome.com).

Color settings: Choose how to color the word cloud. Three options are available. If “Color theme” is chosen, you can pick a color theme from the dropdown list of options and the range value that determines which color(s) in the theme are used for which values in the word cloud. If “Text color” is chosen, you can pick a specific color from the colors available on the system. If “Mask color” is chosen, the colors of the words in the word cloud will match the colors used in the original mask image. The image background can be set to a specific color or made transparent.

Font settings: Choose which font family and font size use in the word cloud. If the “Allow squeezing” option is selected, new words to be added to the word cloud that cannot fit in the remaining space will be incrementally “squeezed” (reduced in font size) until the fit the space available. This “squeezing” effect will distort the appearance of a word but will usually result in a more aesthetically pleasing result. It is also used when the “Repeat words” option is chosen to fill as much of the remaining space as possible. The “Scaling Factor” setting determines the weighting given to the value (e.g., frequency) of the word to the ranking of that value. At 1.0, only the value is considered. At 0.0, only the rank is considered. The “H(orizontal)/V(ertical) Ratio” setting determines the probability of horizontal words plotted over vertical words. At 1.0, all words will be plotted horizontally. At 0.0, all words will be plotted vertically.

Scratchpad: This is free writing area. If the scratchpad is chosen as the source, the words here (and their frequencies of occurrence) will be used to plot the wordcloud.

SEARCH OPTIONS

Search queries can be composed of full words or word fragments, with or without wildcards. The basic syntax roughly follows the Common Elementary Query Language (CEQL). See https://cwb.sourceforge.io/ceql.php for more details). Searches can be either “case insensitive” (default) or “case sensitive” by activating or deactivating the "Case" search term option. Searches can also be made using full regular expressions by activating the "Regex" option. With "regex" option, each word-level regular expression needs to be separated by whitespace. To make regex expressions case-aware, select the "Case" 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 (e.g., http://www.regular-expressions.info/quickstart.html). AntConc supports Perl regular expressions including Unicode character classes, e.g., \p{Letter}, even though the software is built using Python.

ADVANCED SEARCHES

By clicking on the "Advanced Search" button (available in all tools), more complex searches become possible.

- The "Search Query List" option allows you to import a set of search queries. You can do this in one of three ways: 1) Type each individual search query in the entry box and click "Add"; 2) Drag and drop the list of search queries into the viewer below the entry box; 3) Copy and paste the list of search queries into the viewer below the entry box. When dragging and dropping or pasting your list of queries, each line will be treated as a separate search query. This feature allows you to use a large set of search queries without having to retype them each time. Any search query accepted in the main interface can be used.
- The "Context Search List" option (not available in all tools) allows you to define search queries that must match within a certain context window around the main search term(s). For example, to search for "student" or "students" appearing at least three words to the left or right of the word "university," add

"university" to the "Search Query List", and then add "student" and "students" to the "Context Search List" list. Finally, set the "Context Search List" "Window Span" as "From 3L" and "To 3R".

- The "SQL Search List" option (not available in all tools) allows you to adjust query conditions by applying conditions directly on tables in the database that stores the corpus through a series of 'join' operations. For example, the following entry in the list will join a custom "genres" table in the database (created through the Corpus Manager) with the main "corpus" table, applying a condition that the "genre" column entry for the file must be marked as "academic". The two tables are joined via the common "doc_id" column:

  ["genres", "genre = 'academic'", "doc_id"]

  The format for the list entry should be a JSON array, with three components, "table" + "condition" + "join column"

MENU OPTIONS

Menu options are divided into three groups, "File", "Edit", "Settings" and "Help". The options available in each group will be described below.

<FILE>

- Create Quick Corpus...
  - This option is for quickly creating a temporary corpus. You will be asked to choose the files you want to add to your corpus. Then, the software will create a "temp " corpus using the default settings of the Corpus Manager. Any existing "temp" corpus will be overwritten.
- Open Corpus Manager...
  - This option opens the Corpus Manager, where you can choose prebuilt corpora from the default library, add or delete corpora from a user library, or create custom corpora from raw files. See the Corpus Manager for complete details.
- Clear Tool/All Tools/All Tools and Files
  - These options will reset the interface
- Save Current Tab Results...
  - This option allows results to be exported in a file format. Direct copying and pasting of results from the interface are also possible
- Save Current Tab Datatables...
  - This option allows complete tables of results from the corpus database to be exported to a set of .csv files. All relevant information about results can be found in these files.
- Import Settings From File.../Export Settings To File...
  - These options allow the state of the software to be saved and reloaded at a different time
- Restore Default Settings
  - This option resets the state of the software to when it was first launched. All custom settings are lost.

<EDIT>

- Select All
  - This option selects all results in the results window. The same effect can be achieved using the standard keyboard shortcut for "Select All". See the SHORTCUTS section for more details.
- Copy
  - This option selects any text in the results window. The same effect can be achieved using the standard keyboard shortcut for "Copy". See the SHORTCUTS section for more details.

<SETTINGS>

- Global Settings (applied to all tools in the interface)

    - Colors - decides the main highlight color (e.g., for highlighted words in the File tool)
    - Files - decides how the paths to files are shown. Also, this setting determines which file types are used as defaults in File Open dialogs and file drop options.
    - Fonts - decides the font family, size, and style of the font for the main interface
    - Language Direction - decides how to display results (especially for the KWIC concordance tool) depending on the language direction. For example, choose the default "Left-to-right" option for language such as English. Choose the "Right-to-left" option for languages such as Arabic.
    - Searches - lists all wildcard available in the system (note that these cannot be edited)
    - Statistics - decides how values are displayed in the interface
    - Restore Settings - decides if the settings will be automatically saved and restored when AntConc is restarted
    - Tags - decides how tags are displayed in the interface
    - Tool Filters - decides if only words in the selected file will be shown or hidden in the respective tools. When the "Hide words in file" option is chosen, the selected file serves as a "stop list".
- Tool Settings
    - KWIC
        - Display Options - decides the colors used to highlight the sort order
        - Other Options
            - Choose to show or hide the file names in the display
            - Choose to show or hide the search term in the display. This option is useful for allowing instructors to quiz students on possible words to fit the gap.
    - Plot
        - View Style - decides which view to use (table/graphic or graphic)
        - Display Options - decides how results are displayed
        - Statistics - decides the parameters for determining the dispersion measure
        - Other Options - decides various parameters for sizing/displaying the plot graphs
    - Cluster
        - Display Options - decides what information is shown in the results window
        - Filter Options - decides if clusters can only span cross whitespace boundaries or can include other characters (e.g., punctuation)
    - N-Gram
        - Display Options - decides what information is shown in the results window
        - Filter Options - decides if clusters can only span cross whitespace boundaries or can include other characters (e.g., punctuation)
    - Collocate
        - Likelihood Measure + Threshold
            - Choose the statistic and cut-off point (threshold) for inclusion of words in the collocates list. Words below the cut-off-point are deemed to appear frequently together with the query term by chance.
        - Effect Size Measure + Threshold
            - Choose the statistic used determine the strength of relationship between the query term and collocate and a cut-off point (threshold) for inclusion of words that meet the minimum effect size.
    - Word
        - List Type - decides what type of list is shown. The "Word List" option will show the frequency of words in the corpus. The "Headword List" option will conflate the frequencies of individual words in the corpus under a grouping headword (if available). The family members for each headword will be displayed with their frequencies of occurrence appearing in parentheses.
        - Display Options - decides what information is shown in the results window
    - Keywords

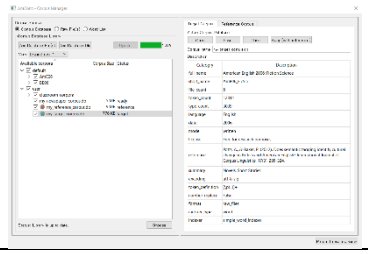
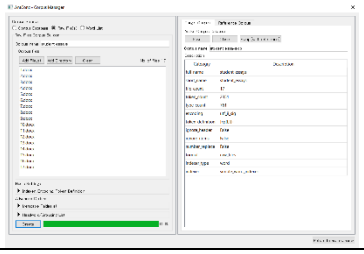
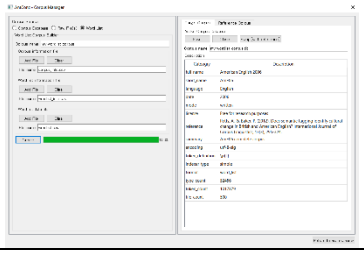
- Display Options - decides what information is shown in the results window
- Negative Keywords - decides to show words in the target corpus that appear unusually *infrequently* in the target corpus compared with the target corpus.
- Likelihood Measure + Threshold - decides the statistic and cut-off point for inclusion of words in the keyword list. Words below this cut-off-point are deemed to appear frequently in the target corpus compared with the reference corpus by chance.
- Effect Size Measure + Threshold - decides the measure used determine the strength keyness and a cut-off point for inclusion of words that meet the minimum effect size.

[Appropriate effect size measures are still being debated in the field, so the default setting is to show all values for this measure. With the default settings, keywords are ranked according to their likelihood measure scores. This equates to ranking keywords according to p-values, which raises several questions/problems. However, it is the current standard in the field and results tend to show that ranking by likelihood leads to more intrinsically intuitive results than those generated when an effect size measure is used. The current selection of likelihood measures and effect size measures are inspired by the work of Andrew Hardie of Lancaster University.]

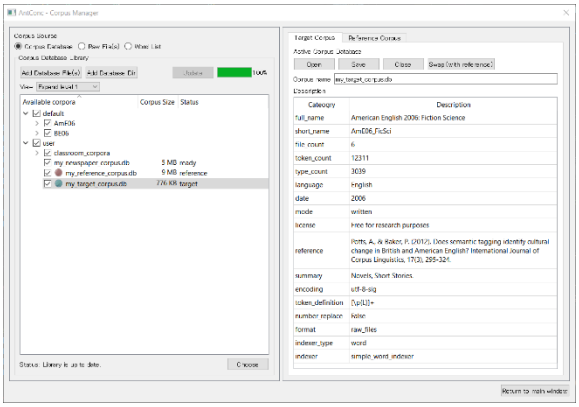
< HELP >

- Show Help Page
  - This option shows the current help guide as a PDF file.
- Show License
  - This option shows the license agreement that you agree to when using the software.
- Show Version History
  - This option shows the complete history of releases, detailing new features, bug fixes, and major updates.
- About AntConc
  - This option shows the release version, release date, copyright information, and acknowledgments for the software.

Corpus Manager

Choosing/Saving a pre-built corpus	Building/Saving a corpus from raw files	Building/Saving a corpus from a word list
		

The Corpus Manager is a multi-purpose tool used to load and save pre-built corpus databases, create and save a new corpus from raw files, or create and save a new corpus from a word list. The three different scenarios are explained below.



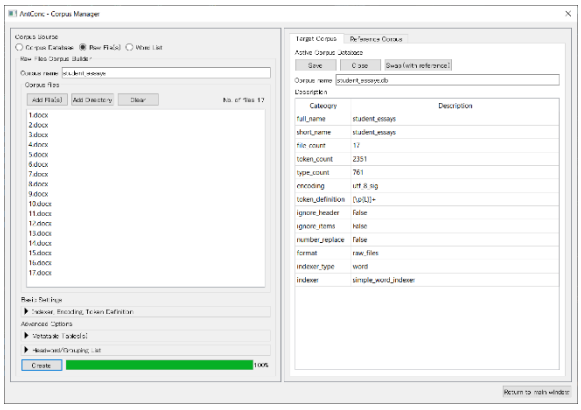
Choosing/Saving a pre-built corpus database

Choose the "Corpus Database" corpus option.

1. If you click this option, a list of pre-built corpus databases available in the "Corpus Library" of AntConc is shown in the left windowpane in a tree layout.
2. The list shows all “default” corpora that are available with AntConc via an online repository. Pre-loaded corpora will be marked with a checked box. Others can be downloaded from the repository, by checking the box next to their name and clicking “Update”.
3. The list also shows all “user” corpora that are stored when you create your own corpora or add a corpus database to the library directly.
4. You can add pre-built corpora directly to the "Corpus Library" list using the "Add File(s)" or "Add Directory" options. They will appear under “user”. Once you have added the files, you can choose one of these by selecting it and then clicking "Choose" or double clicking on the list item. Any corpora built from raw files or a word list (see below) will be shown under “user”
5. Choose one of the “default” or “user” corpus databases by selecting it and then clicking "Choose" or double clicking on the list item. A description of the chosen corpus will be shown in the right windowpane.
6. Unchecking the box next to the name of a corpus and clicking “Update” will remove the corpus from the library.
7. Pre-built corpora can be accessed directly (without copying them to the “Corpus Library” by clicking on the "Open" button in the right windowpane and choosing the corpus file.
8. Any chosen corpus can be saved to a new location by clicking on the "Save" button in the right windowpane and saving the file.

Click the "Return to Main Window" button at the bottom right of the Corpus Manager to return to the main window and start using the corpus. The corpus files will be shown in the left pane of the main window.

Building/Saving a corpus from raw files



Choose the "Raw Files" corpus option. Several 'builder options' will appear in the left windowpane. Follow the steps below:

1. Choose a name for your custom corpus. A default name is provided.
2. Choose the files to be included in your corpus
    a. Use the "Add File(s)" or "Add Directory" options to choose your raw files. You can choose plain TEXT (.txt), WORD (.docx) and PDF (.pdf) files.
3. Adjust one or more of the basic settings as necessary (OPTIONAL)
    a. Decide the indexer used to process the raw files.
        i. For simple files with no annotation or part-of-speech (POS) tagging, the "simple_word_indexer" will work well.
        ii. For simple files that have been part-of-speech (POS) tagged, the "simple_word_pos_headword_indexer " will work well.
        iii. For simple files that have been part-of-speech (POS) tagged using the Biber Tagger, the " simple_word_bibertag_indexer " will work well

iv. For other files with more complex structures, different indexers will become available over time.

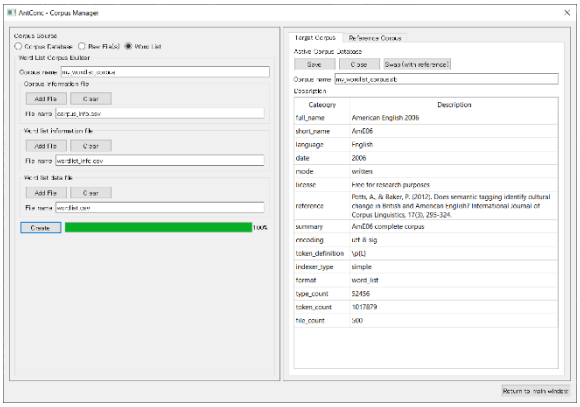
b. Choose the character encoding of the files
   i. The default option (UTF-8) is the standard in the field. Many other encodings are also available in the "Other" options.
      - If you are unsure of what encoding to use, it is recommended to initially choose the default UTF-8 option. Later, if you see an encoding error when trying to process your files or you find that your files appear corrupted in the various tool displays, it probably means that the encoding is wrong, and you should determine the correct encoding to use. For Word (.docx) and PDF (.pdf) files, the default encoding should generally be fine.
   ii. For more information on the Unicode standards see:
      - http://www.cs.tut.fi/~jkorpela/unicode/guide.html
      - http://www.unicode.org/
      - http://www.unicode.org/Public/5.0.0/ucd/UCD.html
      - http://www.unicode.org/Public/UNIDATA/PropList.txt
      - http://www.unicode.org/charts/

c. Decide the definition of a token (word) in the corpus.
   i. In some cases, you may only want to include tokens (words) comprised of the letters a-zA-Z in your corpus, whereas other times, you might want to include tokens comprised of letters, numbers, apostrophes etc. The token definition determines what tokens your corpus is comprised of.
   ii. If you click on the "Show Token Definition Settings" button, a new window will open where you can choose your definition. AntConc offers three ways to choose a token definition: "Character Classes", "User-Defined Characters", and "User-Defined Regex":
      1. "Character Classes": This is the default option and is the most comprehensive. Click the various options to add characters to the definition. These classes are fully Unicode compliant, meaning that they can handle data in any language, including all European languages and Asian languages. For example, the default option "Letters" refers to 'letters' in the broadest sense, including all English letters (a to z, A to Z) and also all Japanese and Chinese 'letter' characters.
      2. " User-Defined Characters": This is a simple option whereby all characters you type in the text edit box will be included in the final token definition.
      3. " User-Defined Regex": This is another simple option whereby all characters that match the given regex (regular expression) will be included in the final token definition.
   iii. If you activate the "Ignore header" option, you can choose a starting and ending tag for your corpus files headers. Any text between these two tags will be ignored when the corpus is created.
   iv. If you activate the "Ignore footer" option, you can choose a starting tag for your corpus files footers. The footer tag and any text appearing after the tag be ignored when the corpus is created.
   v. If you activate the "Ignore non-embedded tags" option, you can choose a starting and ending tag for elements in your corpus files that you want to ignore. Any text between these two tags will be ignored when the corpus is created.
   vi. If you activate the "Ignore embedded tags" option, you can choose a tag maker for token elements in your corpus files that you want to ignore. Any text appearing after the tag marker will be ignored when the corpus is created.
   vii. Using the "Token Testing Area", you can test your token definition by typing or copy/pasting a text into the left-hand text box, clicking the "Test" button, and checking in the right-hand text box to see what tokens will be generated.

viii. After defining your corpus token definition, click "Apply"

4. Adjust one or more of the advanced settings as necessary (OPTIONAL)
    a. Choose metadata tables (if available)
        i. If you click on "Add File(s)" or "Add Directory", you can choose optional metadata tables that will be stored as SQLite database tables together with your raw corpus data. The information in these metadata tables must be aligned with the column names used in the existing tables of the corpus. To understand the default table structure, open the corpus database in an SQLite database reader (e.g., https://sqlitebrowser.org/) and view the different tables.
        ii. Once the corpus is built, you will be able to form search queries on the main corpus using values in these tables as conditional elements.
    b. Choose a headword/grouping list (if available)
        i. If you click on "Add File", you can choose an optional headword/grouping list that will be used to map words to headwords or grouping categories. Existing headwords (e.g., those generated by a POS tagger) will be overwritten by these headwords.
        ii. Once the corpus is built, you will be able to form search queries on the main corpus using these headword/grouping terms as conditional elements.
5. Create the corpus
    a. To complete the corpus building process, click the "Create" button.
    b. Once the corpus is created, a basic description will be displayed in the right windowpane.
    c. The corpus will be available in the "Corpus Library" and can be saved to another location by clicking on "Save" button next to the “Active Corpus Database” label.

Building/Saving a corpus from a word list



Choose the " Word List" corpus option. Several 'builder options' will appear on the top-right of the window. Follow the steps below:

1. Choose a name for your custom corpus. A default name is provided.
2. Under "Corpus information file", click the "Add File" button to choose a "Corpus information file" in CSV or TSV format that matches the "corpus_info" table structure used by AntConc.
3. Under "Word list information file", click the "Add File" button to choose a "Word list information file" in CSV or TSV format that matches the "wordlist _info" table structure used by AntConc.
4. Under "Word list data file", click the "Add File" button to choose a "Word list data file" in CSV or TSV format that matches the "wordlist" table structure used by AntConc.
   [To understand the various formats required, load one of the existing pre-built corpora, generate a word list, and then save the database tables via the File menu "Save Current Tab Datatables" option. In the saved .zip file, you will find examples of the table formats that you need to use. Alternatively, open the corpus database in an SQLite database reader (e.g., https://sqlitebrowser.org/) and view the different tables. An SQLite database reader should allow you to export the tables as a CSV/TSV files that you can use as templates for your own wordlist corpus creation.]
6. Create the corpus
    a. To complete the corpus building process, click the "Create" button.
    b. Once the corpus is created, a basic description will be displayed in the right windowpane.
    c. The corpus will be available in the "Corpus Library" and can be saved to another location by clicking on "Save" button next to the “Active Corpus Database” label.

[Note that a wordlist corpus can only be used as a reference corpus for use when generating keyword lists. Using a wordlist-based corpus in other tools will result in an error.]

SHORTCUTS

Here is a list of useful shortcuts (including some of the useful standard shortcuts on the operating system).

- CTRL/COMMAND + TAB: Toggles clockwise through the different tools in the tab bar.
- ALT + Tool Number: Selects a specific tool (e.g., ALT+1 for the KWIC Tool, ALT+2 for the Plot tool).
- SHIFT + CTRL/COMMAND + TAB: Toggles anti-clockwise through the different tools in the tab bar.
- CTRL/COMMAND + C: Copies the currently selected text.
- CTRL/COMMAND + A: Selects all text in the window.
- F4 (Win): Reveals the complete list of options in a 'combobox' widget (e.g., the search history in the search query box).
- ARROW KEYS: For any 'combobox' widgets (e.g., the KWIC search query box) or 'spinbox' widgets (e.g., the KWIC context size), the 'UP' and 'DOWN' arrow keys on the keyboard can be used to change the value of the option.
- CTRL/COMMAND + O: Opens the "Corpus Manager".
- CTRL/COMMAND + G: Opens the "Global Settings".
- CTRL/COMMAND + T: Opens the "Tool Settings".
- CTRL/COMMAND + H: Toggles the view "(H)ide" of "Show" setting for the KWIC tool file name column.
- CTRL/COMMAND + '+': Zooms in the "Plot Tool" display.
- CTRL/COMMAND + '-': Zooms out the "Plot Tool" display.
- CTRL/COMMAND + F: Searches for the next hit in the "File Tool" display.
- SHIFT + CTRL/COMMAND + F: Searches for the previous hit in the "File Tool" display.

NOTES

- If you have any suggestions for improving the software or notice any bugs, please post them in the AntConc Discussion Group (https://groups.google.com/g/antconc). Indeed, many of the improvements and updates made to the software have been due to the comments of users around the world, for which I am very grateful. The AntConc Discussion Group is also a good place to discuss how you are using the software and any challenges that your face.
- If you find the software useful in your research, teaching, or learning, you may consider making a small donation to support the future development of this tool. A link to the donation page can be found here: https://www.laurenceanthony.net/software/antconc/
- You may also be interested in becoming an AntConc patron. Depending on the level of support, this option will give you priority support with direct access to the developer (Laurence Anthony), and various other benefits. A link to the donation page can be found here
- https://www.patreon.com/antlab

CITING/REFERENCING ANTCONC

Use the following method to cite and reference AntConc according to the APA style guide:

Anthony, L. (YEAR OF RELEASE). AntConc (Version VERSION NUMBER) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/

For example, if you download AntConc 4.0.0, which was released in 2021, you will cite/reference it as follows: Anthony, L. (2021). AntConc (Version 4.0.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/

ACKNOWLEDGEMENTS

I would like to say thank you to the users of *AntConc* who have taken the trouble to post feedback on the software and make suggestions for improvements and/or changes. A very special thank you goes to all those who have very generously supported the project either through single donations via PayPal or becoming a Patreon supporter. A complete list of individual acknowledgments can be found in the Help menu - "About AntConc" menu.

The development of *AntConc* has been supported by a Japan Society for Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (C): No. 23501115, a Japan Society for Promotion of Science (JSPS) Grant-in-Aid for Young Scientists (B): No. 18700658, a Japan Society for Promotion of Science (JSPS) Grant-in-Aid for Young Scientists (B): No. 16700573, and a WASEDA University Grant for Special Research Projects: No. 2004B-861.

KNOWN ISSUES

None at present.