



AntConc (Windows, MacOS, 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

VERSION HISTORY

4.0.0

This is a major upgrade. The code for AntConc (originally written in Perl) has been completely rewritten from the ground up in Python. For a complete understanding of the new version, please read the help file.

Some of the important new features (many more are not listed here):

1. All processing is handled in a SQLite database. This means all processing speeds should now scale well. By directly editing the SQLite database, users should be able to add extra data, extra functionality, and tweak or in some cases greatly improve the performance of the system (by adding custom indexes on their corpora). The database tables can be easily saved, allowed the data generated by AntConc to be easily used with other tools. Similarly, database tables created by other tools, can be loaded into AntConc for processing.
2. The loading and saving of corpora are now managed through a "corpus manager" tool. Prebuilt corpora stored in SQLite databases can be loaded (as target or reference corpora) or saved directly through the interface. Raw files can also be processed and saved in an SQLite database format through the interface allowed the data to be portable and rapidly accessed at future times. The "corpus manager" can also be used to load word lists (when full corpora are not available), and associate lemma/group headwords to raw words for later analysis.
3. Raw corpus files in the form of plain text (.txt), Word (.docx), and PDF (.pdf) files can be processed directly and converted into a single SQLite database.
4. All results are now displayed in a paged format, which can be adjusted to show a small number of results at a time, or all the results in a single page. For very large corpora, paging results in small numbers hugely improves the performance of the software.
5. Results for frequency and range are shown in all tools together with normalized values where appropriate. This should lead to easy comparisons of results across corpora.
6. The Key-Word-In-Context (KWIC) Concordancer tool, now re-named as "KWIC", can sort concordances alphabetically (the current standard in the field) or more usefully by the frequency of patterns according to the sort parameters. This new display option will be referred to as a "KWIC pattern display" in all future documentation. The table rows can also be easily copied and pasted directly into other tools (e.g., Word/Excel) with the color highlighting intact.
7. The Concordance Plot tool (now re-named as "Plot") shows various statistical measures in addition to the existing plot display in a more conventional table format. A purely graphical view of the plot display can also be chosen via the tool menu options. The purely graphical view can be generated very much quicker than the table/plot view and is recommended for plotting results for a large number of files and hits. Plotted results can also be overlaid with new results (in a different color).
8. The File View tool (now re-named as "File") is largely unchanged, but results can now be easily copied and pasted directly into other tools (e.g., Word/Excel) with the color highlighting intact.
9. The Cluster/N-grams tool have now been split into separate tabs. The Cluster tool now only allows one cluster size to be set (instead of a range of values as in previous versions of the software). Similarly, the N-gram tool now only allows one n-gram size to be set.
10. The Collocate is largely the same as in previous versions although performance has been greatly improved.
11. The Word List and Keyword List tool are largely the same as in previous versions although, again, performance has been greatly improved. In addition, when searching within the lists, the lists are now filtered to display only the searched items. This feature makes viewing the results substantially easier than in previous versions.

3.5.9

This is a minor upgrade with a single bug fix.

Bug fixes:

1. Fixed a bug in the File View tool that caused highlighting and results to be mismatched when the context search option was used. This bug also impacted results when linked from other tools such as the Concordancer.

3.5.8

This is a minor upgrade with several bug fixes.

Bug fixes:

1. Fixed the label of the two-level statistic "MI + Log-Likelihood ($p < 0.05$)" to correctly show the p value to be less than 0.05. Originally it suggested $p > 0.05$.
2. Re-introduced the display of frequency values of lemma family members in the Word List tool when the lemma option is activated.
3. Removed the appearance of irrelevant index data values after file names in the output of the KWIC concordancer tool.

3.5.7

This is a minor upgrade with several bug fixes.

Bug fixes:

1. Scroll bars in the Word List tool would only be properly shown if the cursor was momentarily moved into the results window or the scroll bar was directly moved. This is now fixed.
2. The Word List tool reported the total number of types and tokens regardless of whether or not a lemma list was used. This is now fixed to properly show the total lemma types when a lemma list is used.
3. A bug dating back to the AntConc 3.4.x version s that caused KWIC concordancer results to misalign with those in the File View tool when a search pattern produced results that spanned line breaks has now been fixed.

3.5.6

This is a minor upgrade with several important bug fixes.

Bug fixes:

1. A bug introduced in version 3.5.4 caused the loading of lemma lists to print out the lemma in the background dramatically slowing the loading process. This is now fixed.
2. A long time bug that caused the tab embedded files to be rendered incorrectly on Mac OSX has now been fixed. I only became aware of the issue very recently.
3. A long time bug that caused files that had only character returns marking line breaks to load on a single line has now been fixed. Again, I only became aware of the issue very recently.
4. The routine to automatically remove BOMs from UTF-8 encoded files has been improved to be more reliable.
5. A bug that caused index numbers to appear in the file column of cloned KWIC concordance windows has now been fixed. The index numbers no longer appear.
6. A bug that caused the results to be automatically sorted on first generation (regardless of the sort options) has now been fixed. It was also noticed that in version 3.5.x it was not possible to view the result in the order they appeared in the file (because the first sort option was also activated). This has been fixed by allowing all sort levels to be activated and deactivated. When all the sort options are all deactivate, the results will be in the order that the hits appear in the files.

3.5.5

This is a minor upgrade with fixes to small but important bugs that affected results in the KWIC concordance tool and were introduced in version 3.5.4.

Bug fixes:

1. A bug introduced in version 3.5.4 caused the KWIC concordance window to only a limited number of characters to the right of the search term and also not show characters that appeared across line boundaries. This is now fixed.
2. A bug introduced in version 3.5.4 caused the KWIC concordance window to skip edge case results when the same search term appeared repeatedly close to itself. This is now fixed.

3.5.4

This is a minor upgrade with a single but important bug fix that hugely improves the performance when working with large files.

Bug fixes:

1. The new way to calculate KWIC concordance lines introduced in version 3.5.x was inadvertently very slow when dealing with single, very large files. This has now been fixed. However, it has also been noticed that the performance of searches is 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).

3.5.3

This is a minor upgrade with a single important bug fix.

Bug fixes:

1. The Advanced Search window was now showing up correctly on Macintosh OS X computers. This is now fixed.

3.5.2

This is a minor upgrade with a single important bug fix.

Bug fixes:

1. The Word List Tool search function was not showing any highlighted results. This is now fixed.

3.5.1

This is a minor upgrade with a single important bug fix.

Bug fixes:

1. The Advanced Search display on Linux was not appearing correctly. This is now fixed.

3.5.0

This is a minor upgrade with various new features and bug fixes.

New Features:

1. The numbers of rows presented in the Concordance Tool and lines plotted in the Concordance Plot Tool can now be restricted via the "Show Every Nth Row" option.
4. Two new statistics are included in the collocates tool preferences. These are "Log-Likelihood" and a two-level statistic "MI + Log-Likelihood ($p < 0.05$)" that combines the "Log-Likelihood" statistical measure at a threshold of $p > 0.05$ with the MI effect-size measure.
5. The Concordance Plot tool has been improved to show hit numbers within the plots and also allow the plots to be saved to a postscript file. The postscript file can be easily converted to a PDF file through Adobe Acrobat or a similar tool. In addition, the plot tool can now plot hits in a normalized fashion (as in previous versions) or using a relative length to the length of the file (in characters).
6. Shortcuts are now adjusted on Macintosh OS X systems to more closely match normal keys.
7. The Word List Tool can now accept a file containing a word list.
8. The Keyword List Tool now offers a clearer set of keyword statistic critical values.

9. The Keyword List Tool now offers a choice of effect size measure by which to rank the results.
10. The Keyword List Tool now offers 4-term Log-Likelihood keyness measure (set as default), which is considered to be the current best practice for Log-Likelihood keyness calculations.
11. The Word List Tool and Keyword Tool now use a backend database to do all processing. This speeds up the processing.
12. The encoding options now include an Auto-Detect option. This will be very successful at detecting Unicode encodings – e.g. UTF-8, UTF-16LE, UTF-16BE, and recent improvements means that it also has very good success at detecting many other legacy encodings. The detection engine is the same as that used in the FireFox web browser.

Bug fixes:

1. Many spinboxes were set to maximum values of 99 or 100. They are now set at a max value of 999999.
2. The zoom function in the plot tool now only expands the plots in the horizontal direction. This allows users to more easily see tightly packed dispersions.
3. Deleting lines in the concordance display will now correctly delete the same lines in the plot display.
4. Save file menus now correctly report the file type to be used helping the user to select a suitable file more easily.
5. Saving of the results in the File Tool produced a blank file. This is now fixed.
6. When opting to replace whitespace with a character in the global settings wildcards menu, the result was always to replace whitespace with the number 1. This is now fixed.
7. Early dev versions had a bug that caused the Plot Length Options menu to show the normalized length and relative length options as checkboxes instead of radioboxes. These are now fixed.
8. Although not strictly a bug, regular expression (regex) searches now use back-references starting with \1 (instead of \2 as in previous versions).

3.4.4

This is a minor upgrade addressing one bug.

Bug fixes:

1. When the "Hide tags" or "Hide tags (Search in Conc/Plot/File View)" global setting tag options were selected, the File View would continue to show headers in certain cases, i.e., when the header extended beyond a single line. This would cause the strange effects such as the concordance/concordance plot lines and linking to become out of sync. This has been fixed.

3.4.3

This is a minor upgrade addressing two bugs.

Bug fixes:

1. When searching using the "Hide Tags (Search in Conc/Plot/File View)" option in the File View tool, if the target file did not contain both non-embedded and embedded tags, the program would freeze. This is now fixed. I am not sure how long this bug has existed. The problem may have been introduced due to a change in the core programming language, but I cannot be sure.
2. The Windows "Delete" and "Insert" shortcut keys on Windows systems are now mapped to the "Delete" and "Shift-Delete" keys on Macintosh OS X systems.

3.4.2

This is a minor upgrade addressing a refresh issue that was noted on Macintosh OS X 10.9 (Mavericks) operating systems.

Bug fixes:

1. Due to changes introduced to the graphics rendering on Macintosh OS X 10.9 (Mavericks), AntConc did not refresh correctly after certain actions making the screen look to freeze. The earlier workaround was click outside of the AntConc window and back into the window to force a refresh. This new version of AntConc fixes the problem at the code level and so no refresh issues should be seen in the future. The workaround has involved creating a standard build of the Perl programming language based on Citrus Perl (5.16.0) and the latest Tcl graphics engine. The introduction of the latest versions of these tools may introduce other very small changes in results due to the use of the latest Unicode standard, but these are unlikely to affect normal users.

3.4.1

This is a minor upgrade addressing several issues noted with the release of 3.4.0.

Bug fixes:

1. On Macintosh and Linux systems, I have heard reports that the advanced search window sometimes does not appear correctly when the "Advanced Search" button is clicked. To address this, I have set the window size explicitly (instead of calculating the size from the internal widgets) and forced a refresh of the screen. This addresses the problem on the systems that I have checked (Macintosh 10.6.8, 10.7.2 and Linux Mint 10). I hope it fixes the problem on other systems. This is difficult to test so please let me know if you continue to find a problem.
2. If no search term is inputted into the search box, an error that "No search term has been entered" is reported even when terms have been entered using a file based search. This is now fixed.
3. In 3.4.0, when the program was first started, the collocates display would incorrectly show the span to be 5L to 5R, when it was in fact 1L to 1R. Although changing the span setting would rectify the problem, on first load, the results would appear to be incorrect. In this new version, the display continues to show the span to be 5L to 5R, but the results now match the setting.
4. In all versions of AntConc, the results produced when using context horizons (e.g. in the advanced search of the Concordance tool) are dependent on the window span (e.g. set in the main windows of the Concordance Tool). If the span of the horizons goes beyond the window span, spurious results can be generated. In 3.4.0, due to the nature of results being stored in a set of columns in database, if the left horizon span went beyond the number of columns estimated by the window span, no hits would be reported. In some ways, this was a good thing as it showed that something is wrong with the span settings. However, the results no longer matched those in previous versions of AntConc that simply ignored the mismatch and produced results for the largest horizon span possible. I have now edited the storage routine so that 3.4.1 also ignores the mismatch and creates results for the largest column size possible. I am still unsure if this is best approach to take. I am considering if a proper error or warning should be reported when a mismatch between the context horizon span and the window span arises.
5. Although not strictly a bug, due to changes in the way the Clusters/N-Grams tool treated line breaks introduced in 3.4.0, results stopped matching those from earlier versions of AntConc (e.g. 3.3.5). In this version, I have introduced an option in the Clusters/N-Grams preferences to either replace line breaks with a space or other character (as happens with the Concordance Tool - the default), or leave line breaks as is. If line breaks are left as is, the results will match those from previous versions of the software.
6. Although not strictly a bug, on startup with a user settings file, AntConc would not remember the size or position of the main window and instead show it at the default size and position. This has been fixed.
7. When AntConc is launched on a Windows system with a magnified view (e.g. 125%), some of the settings options were hidden from view. To compensate for this feature of Windows, I have redesigned the settings windows so that all options are viewable when first launched in magnified view at 125%. Viewing at 150% is still likely to hide certain options. I have also allowed the settings windows to be resized so that even at 150%, it should still be possible to resize the settings window to see the options.
8. The way the program reads in files has been changed to avoid a strange bug that occurred when reading in some Unicode files that resulted in the last character on a line being misread on very rare occasions.

The way file names are displayed in the program has also been changed to avoid a potentially similar problem (although I have heard of no reports on this to date).

9. The "Range" heading of cloned Clusters/N-Grams tool results did not show correctly. This has been fixed.
10. The "Stat" heading of the cloned Collocates tool results was shown as "Prob". This now appears correctly.

3.4.0

This is a semi-major upgrade fixing a few small bugs and adding several important features to improve the output display of the Concordancer and Plot tools.

New Features:

1. The Concordancer tool output now correctly centers the search terms regardless of language. This is especially useful for Asian languages such as Japanese, Chinese, and Korean that have two-byte character widths. Note that this is achieved by using two columns in the 'KWIC' pane. If you click CTRL-A in the pane, only one of these hidden columns will be selected. Use ALT-A to select all columns.
2. Line breaks in the Concordance Tool output can now be replaced by a user-defined character (default being a single space). Again, this is especially useful for Asian languages because traditional replacement by a single space might inappropriately split words that cross line-break boundaries.
3. The Collocates Tool no longer includes the center word in the collocation table. This results in more meaningful collocate values.
4. Not exactly a feature, but Concordance Tool searches are now only centered on completion rather than after finding the first line. This is generally a more pleasant experience.
5. Whitespace options in the wildcards menu have now been extended.
6. A database now runs in the background and is used to store KWIC and Plot concordance results. This alleviates out of memory problems with earlier versions and allows much faster sorting of the KWIC. The new database model also improves the performance of general KWIC and Plot searches.
7. In the Advanced Search, if you add multiple context words, AntConc will look for *any* of these in surrounding context of the search term rather than *all* of them (as in earlier versions). This change has been implemented as a result of many requests for this function. I hope it does not inconvenience too many people.
8. In the Collocates tool, if you click on a word, it will search for that word in the Concordance Tool but restrict results to only those where the original search term is in the neighborhood of the collocate as set by the Collocates tool. Again, this change has been implemented as a result of many requests for this function.
9. Colors for the highlight and ordered words in the KWIC display have been changed to be a little easier to see.
10. The font size for the main results is now a little larger.
11. Mismatches in results across tools (e.g. different frequency counts for the Concordance, Plot, File View and Clusters/N-grams tools) due to the different ways in which tools handle line breaks and spacing have been fixed. In particular, by changing the default processing to replace line breaks with spaces, the frequencies reported in the Clusters/N-grams tool will now match the Concordance tool hits.
12. All searches are now performed to find the maximum number of hits for a particular search. This means that a search for "aa" in a string "aaa" will now produce two hits instead of one (as in earlier versions of AntConc). For most 'word' based searches, very few differences will be noticed. However, results can be quite strikingly different for string-based searches using wildcards. The new way of doing searches is intuitively more sensible than the previous search approach. However, many other software programs adopt the previous approach so you should expect some differences when comparing results.
13. Generic line break characters representing line breaks on different systems (e.g. `\r\n` for Windows and `\n` for Linux/OS X) are now used for all file processing.

Bug fixes:

1. The Collocates Tool sometimes use to report values of "-1" (when the only collocates of the center word was itself) and "-2" (when the collocate value was not able to be calculated due to an out-of-sync word list). The first case ("-1") no longer exists (see feature 3 above). In the second case ("-2"), the value has now been replaced with "0" to avoid confusion.
2. In the Word List Tool, Alt-A now correctly selects all the panes including the Lemma list.
3. When no corpus file names were selected in the left panel, sometimes the "Close file" menu option would delete the first file name in the list instead of reporting that no file was selected. This has been fixed.
4. After changing the size of the Search Entry font option in the global settings, the new font size would not be reflected in the sample display. This is now fixed.
5. If the "invert sort" option was activated in any tool with this feature, an "Array XXX" error was produced in the results window. I apologize for this very inconvenient bug.
6. Several other minor bugs including one that sometimes led to tags not being hidden properly after applying tag settings have been fixed.

3.3.5

This is a minor upgrade fixing a few small bugs. One very minor new feature has been added.

New Features:

1. The "Order by Stat" sort option is now set as the default option in the Collocates Tool.

Bug fixes:

1. Fixes a problem that caused the window span controller values to become out of sync if the "from" controller changed the "to" controller or vice versa.
2. Fixes a problem that caused the window 'same' checkbox controller to not work properly.
3. Fixes a problem that caused the scrollbars to not show correctly in the cloned view of the Concordancer Tool window.
4. Fixes a problem that caused the background color of the global-settings->character encoding options to appear white on a Mac OS X system. This made the options impossible to read because the OS X native foreground color is also set to white, which is different from that on Windows and Linux computers.

3.3.4

This is a minor upgrade fixing a single bug. No new features have been added.

Bug fixes:

1. Sort by word end in the Word List tool produced a list in reverse order. This is now fixed.

3.3.3

This is a minor upgrade fixing a couple of bugs. No new features have been added.

Bug fixes:

1. A tab delimiter character incorrectly appeared in the KWIC concordance display. This is now fixed.
2. If multi-character delimiters were inputted, the highlighting became misaligned. This is also fixed.
3. A bug causing the software to keep reopening after clicking the close button (top right of screen on Windows) was fixed. The problem only appeared after importing user settings files or reloading the default settings.

3.3.2

This is a semi-major upgrade that addresses an important bug in version 3.3.1. I've also added a few small but useful features.

New Features:

4. Performance has been greatly improved (> 10 times) for all but the Concordance and Plot tools. I hope to improve the performance of these tools from now.
5. It is now possible to append user-defined characters to the standard Unicode character classes.

6. Users can now choose to ignore or not ignore white space in the search box. The default is set to ignore white space
7. Users can now choose which delimiter character to use in the Concordance tool.
8. The "treat all data as lowercase" option is now set as default on all tools except the Concordance and Plot tool. I hope this is a much more sensible default setting.
9. Files created on different platforms (Windows, OS X, Linux) can have different end-of-line characters. These are now treated correctly regardless of which platform *AntConc* is launched on.

Bug fixes:

1. Some user token definitions caused the output to display strange results that appeared to ignore any word boundaries. This problem is now fixed.

3.3.1

This is a minor upgrade that addresses a few small bugs version 3.3.0.

Bug fixes:

1. All windows now resize correctly to show widgets even when the 'screen display' setting on Microsoft Windows is set to 125%. In 3.3.0, an enlarged display caused some setting areas to be cut.
2. Windows are now centered correctly regardless of the screen display' setting on Microsoft Windows. Also, there is no flicker of windows moving from the top left of the screen to the center as they are positioned.
3. The checkbox for Level 1 sorting now includes a greyed out check mark.
4. Underline marks that appeared under list items (e.g. corpus files in the main window) have now been removed.
5. A few bold headers have now been corrected to normal typeface.
6. Word and case checkbuttons in the main window are now correctly disabled if the Regex option is marked as active in a user settings file.

3.3.0

This is a major upgrade introducing many new features and addressing bugs that appeared in previous 3.2.x versions.

New Features:

1. Now *AntConc* runs completely natively on all platforms. This is especially important for Macintosh OS X users as it means no X11 installation is required. Also, the interface is considerably improved over previous 3.2.x versions most spectacularly on Macintosh OS X but also on Windows and Linux systems.
2. The "range" information for a word has now been added to the Clusters Tool. This allows 'lexical bundles' of a certain minimum range to be generated.
3. The full path of a corpus file is now shown by hovering the cursor over the file name in the main window.
4. Loading of lemma lists is now greatly simplified.
5. Loading of reference corpus lists are now greatly simplified.
6. When loading reference corpus lists, checks are now made to see if the format is correct.
7. The interface has been simplified a little by removing little-used buttons, in particular, the "Reset" button and "Exit" button on each tool tab. Also the ambiguous "Save Window" button has now been renamed "Close Results" to more accurately reflect its action.
8. Hit and frequency counts now appear above the results window in all tools.
9. The Clusters tool has now been renamed "Clusters/N-Grams" to more accurately reflect its functions.
10. The actions associated with spinbox arrows have now been reversed. This is perhaps a more natural way for them to work.
11. The term "Language Encoding" used in 3.2.x versions has now been corrected to "Character Encoding".
12. Various default settings have been changed to reflect modern corpus linguists' preferences. In particular, the default character encoding has been changed from Latin 1 to UTF-8.
13. This user file has now been updated to be a little easier to read and understand.
14. After opening or saving files, *AntConc* will now remember the folder location.

Bug fixes:

1. Selection highlighting of corpus files now accurately reflects which files can be viewed in the File View tool.
2. *AntConc* now correctly remembers the positions of window panes after global and tool preferences have been accessed.
3. *AntConc* now correctly remembers which was the last preference window accessed by the user.
4. Numerous interface oddities associated with 3.2.x versions of *AntConc* should now be fixed with the introduction of a native interface for all platforms (see New Features 1 above).
5. File names with a different character encoding the file contents will now appear correctly.

3.2.4

This is a minor upgrade fixing two problems (one only applicable to OS X). No new features have been added.

Bug fixes:

1. Modified the OS X version so that lemma list files will load correctly regardless of the OS version. The bug was more general and possibly caused errors when using any single file open file dialog box, e.g., importing settings files, search list files, and word list files. I only heard one or two reports of problems on OS X, so perhaps the problem was introduced with upgrades to recent versions of OS X. It is difficult to tell.
2. Fixed a bug that caused the total number of lemma entries to not be displayed properly.

3.2.3

This is a major upgrade introducing several new features and addressing bugs that appeared in version 3.2.2.1.

New Features:

1. Massively simplified the engine used for sorting and displaying results. This will allow for improved performance later.
2. Introduced a Concordance preference setting allowing tab spaces to be added around the search hit in the KWIC concordance view. Introducing this means that the toggle shortcut key 'x' to show/hide the KWIC search term has had to be removed. This function can now be accessed via the (fixed) menu option. I also hope to reintroduce this toggle shortcut key later.
3. Introduced the option to use word list(s) of reference corpus files instead of directly processing the raw files. This allows fast generation of keywords and also allows users to generate keyword lists even when the reference corpus is not available (but a word list is). See the Keyword List section for further details.
4. Introduced feature to output the counts of types and tokens for each tool when saving results as a text file.
5. Changed some of the default settings in response to user feedback:
6. Concordance Tool sort levels are set as 1R, 2R, 3R.
7. The Collocates Tool Stats measure is now displayed and calculated.

Bug fixes:

1. The splash screen that appeared when a user settings file was added has been removed. Now, the words "User Settings" appears next to the title of the software at the top of main window, whenever a user settings file is used.
2. The layout of the main window has been improved so that the main work area is maximized for all sizes of main window.
3. Unicode "Marks" can now be added/removed in the token definition settings
4. Occasionally, the last bar of the progress bar remained blank at the end of processing. This has now been fixed.
5. Under tag settings, the "Hide Header Tags" option will now work for headers than span across multiple lines.
6. The main window and all pop-up windows now appear centered in the display.

7. When viewing the global settings preferences, it was possible to accidentally activate the main window by clicking on it. This has been fixed.
8. The list of files in the left pane was often too narrow to view easily. This has been fixed by introducing a two-pane approach, with the left pane showing the files, and the right pane showing the results and controls. The width of both panes can be adjusted with the default size being the minimum size.
9. It was possible to adjust the main window to be smaller than the default size of 800px by 600px. This caused the positioning of some widgets to become misaligned. Now, 800px by 600px has been set as the minimum size of the main window.
10. Various problems experienced when cutting, copying, and pasting text to/from the Search boxes in each tool have been addressed by removing the "Spinbox" widget type and replacing this with a standard "Entry" widget. The history of previous searches can be now accessed using the "Up" (previous) and "Down" (next) arrow keys.
11. After moving the cursor over the corpus files list while using the File View Tool, it should change into a "Pointing Finger" cursor. However, this cursor shape was returned incorrectly back to a standard "Edit" cursor when one of the files was selected. This has now been fixed.
12. Changed foreground to white when *AntConc* reports that the KWIC analysis has finished or has no results. This should be easier to read.
13. Fixed a bug that caused the Concordance preference option "Hide search term in KWIC display" to only show the hit. (The opposite of the stated function!). This is now fixed.
14. Fixed some typos in this Read Me document.
15. Fixed a major bug that caused a warning to always appear when using the T-Score statistical measure with the Collocates Tool. Note that the warning could be ignored as the results would always be correct.

3.2.2.1

This is a minor upgrade addressing bugs that appeared in version 3.2.2.

Bug fixes:

1. The legacy character encodings for Japanese, Chinese, Taiwanese, and Korean were inadvertently left out of the compiled executable for version 3.2.2 meaning that files encoded in these languages would not open correctly in *AntConc*. These have now been added.
2. In the Mac OS X version, the compiled version did not include several important components which are essential for the X11 graphical toolkit to run correctly. These have been added.

3.2.2

This is a minor upgrade addressing several minor bugs that appeared in version 3.2.1. It is also the first version of *AntConc* to be compiled with Perl 5.10.

Bug fixes:

1. Fixed a bug that caused back references to not work correctly when the 'Regex' search option was selected. This now works correctly. However, note that the first back reference should be 2 and not 1 (due to an implicit back reference 1 that holds the entire search result).
2. Fixed error message when the sort by Stat option was selected.

3.2.1

This is a minor upgrade addressing several bugs that appeared in version 3.2.0, as well as introducing a few new features requested by users.

New Features:

1. Better display of long lists of fonts and font sizes in the global options/font menu. Now the lists appear as an easy to navigate list with attached scrollbar.
2. When results windows are saved, the cloned windows now display summary results information.
3. Word range lists can now be used as lemma range lists.
4. New feature allowing tagged data to be searched while remaining hidden. See the tag settings preferences. Pressing CONTROL and the START button (or the ENTER button if the search entry box has the focus) temporarily disables the new feature, allowing the user to switch easily between a 'non-tagged' or 'tagged' display.
5. New options in the tag settings preferences that allow embedded and non-embedded tags to be shown, ignored, or hidden. This enables data of the form of the BROWN and BNC corpora to be processed easily.
6. Improved the updating of the progress bar display, which may also improve the speed of processing in some cases.
7. Improved the images used for icons within the program.

Bug fixes:

1. Fixed bug that caused user defined token definitions containing special regular expression characters from not working properly
2. Fixed bug that caused "Treat all data as lowercase" option to ignore the wordlist range and lemma lists
3. Fixed bug that caused the lemma list "Load" button to not ignore the currently opened file if a new file dialog was opened and then "Cancel" was pressed
4. Fixed bug that caused file searches to not work if the search entry box was blank.

3.2.0

This is a major upgrade with a completely redesigned interface, several new features, and several bug fixes. The new interface follows the basic design used in previous versions, although users should find it 'cleaner' and more intuitive. In particular, all global and tool menu settings have been combined into two groups, where all the related settings can be accessed and adjusted within the same window. This should dramatically improve the usability. All tools now have access to the search engine (including the **Word List Tool** and **Keyword List Tool**) and there is also a new advanced search window that can be used to perform list (file) searches, and searches within a particular context. Due to the nature of the changes, this version will not be compatible with the settings files for previous versions. Another huge change is that this version will run on Macintosh OS X systems.

New Features:

1. Completely redesigned interface
2. Added search and advanced search features to all tools (including the **Word List Tool** and **Keyword List Tool**).
3. Created new list (file) search available in all tools.
4. Created new context search option in all tools except the **Word List Tool** and **Keyword List Tool** where it has no meaning.
5. Busy cursors are used to indicate when very long sorting operations are being carried out (e.g. when sorting large N-gram list results).
6. Case options affecting whether or not data is converted to lower case are now more intuitive. For example, the 'Case' option in the main window now only affects the operation of the search itself and has no impact on the data under observation. Data can be treated as lowercase (for example in Word list tool) by choosing the 'Treat all data as lowercase' under the appropriate category in the 'Tool Preferences' menu.
7. The number of corpus files (and reference corpus files) being analyzed is now displayed.
8. Correct some mistakes in this readme file
9. My name has been removed from the top of the main window! However, please remember that my name is Laurence with a U if you are ever citing me in your research papers!!

10. Now works with Macintosh OSX

Bug fixes:

1. The program no longer crashes when the 'All Values' option is chosen as the threshold value in Keywords.
2. Negative keywords are now highlighted correctly when the 'Show Negative Keywords' option is selected.
3. The KWIC lines are now aligned correctly even when the hit appears near the very start or end of a file.
4. Collocates frequency values are now correctly calculated even when the span extends further left than the start of the file
5. The action of the 'one word only' wildcard is now more intuitive.
6. Some operations (e.g. creating a word list) now do not crash after restoring the default settings and then performing an operation.

Bug fixes (since beta1 version):

1. The program now (correctly) only shows files that generate hits in the Concordance Plot tool.
2. The sort function in the Keywords Tool now works correctly. In previous versions, even when the 'Frequency' option was selected, the sort would be based on Keynes. Also, in some cases inverted sorting did not work.

Bug fixes (since beta2 version):

1. The program now (correctly) hides the various Concordance Tool panes depending on the chosen Display Options. In the earlier beta versions, the options were ignored.
2. The default file type to use when opening directories now works correctly. In previous beta versions, after hitting the apply button, the default file type reverted back to the .txt type.
3. Fixed a bug that prevented the n-grams option in the Clusters Tool from working when the search term entry box was empty.

Bug fixes (since beta3 version):

1. Fixed bug that caused the program to not be able to open files with non-English names correctly if the full-pathname option was selected. There are potentially many problems with non-English filenames, so I recommend that users use English filenames for their corpus files, and also save them under a pathname which only contains English characters.
2. Fixed bug that caused the 'OR' wildcard to not work correctly if a character other than '|' was user defined.

New Features (since beta4 version):

1. Made some small changes so that the program could be more easily ported to Macintosh and Linux platforms.

3.1.303

This is a very minor upgrade with just one change:

Bug fixes:

1. Corrected problem that caused the No. of Hits to not be indicated correctly in the Concordance Plot Tool display when more than one corpus file was being used.

3.1.302

This is a very minor upgrade with the following changes:

Bug fixes:

1. Corrected problem which caused the program to not launch when the path of the default temporary folder on the system contained non-English characters.
2. (Linux only): Corrected problem that caused the Open Dir menu option to not work correctly.

3. (Linux only): Corrected problem that caused font selections to not work correctly.

New Features: Improved speed and memory handling when calculating collocates. Over 10 times faster than in previous versions (including version 3.1.3).

3.1.3

This is a minor upgrade containing an important bug fix that prevented files with non-ASCII filenames being used. There are also some major performance improvements. For example, n-grams will now be processed over 10 times faster on small corpora and many more times faster on larger corpora. A list of all important changes is below:

New Features:

The history feature for search term entries has been changed. I have heard two reports of the 3.1.2 version not starting on computers. Hopefully, this change will allow the program to start correctly on all machines.

The performance of tools such as Collocates, Clusters and N-grams, has been significantly improved. (Over 10 times faster on small corpora and many more times faster on larger corpora.)

The Open Dir option now open files in all sub-directories too.

The program will automatically look for a user defined settings file named "antconc_settings.ant" in the directory where the program is saved. If this file is found, this settings file will be used instead of the default settings. Also a splash window will be displayed when a user settings file is found. If no file is found, the default settings will be used. In this case, no splash file will be shown. This feature allows users to save their setting preferences and use them again without having to load the preferences each time.

Bug Fixes:

4. Files with non-ASCII file names were incorrectly processed preventing them being used. Introduced in version 3.1.2. Fixed.
5. In the N-grams tool, non-word units (e.g. spaces) at the beginnings of lines were treated as words. This caused some n-grams of n-1 size to also appear in the results list. Fixed
6. When "ALL" was selected as the file type option in Open Dir, sub-directories were also displayed even though these could not be opened or processed. Fixed
7. In some tools, non-ASCII filenames were sometimes displayed incorrectly, even after the correct encoding was chosen. Fixed.
8. If 'Exit' was chosen from the File menu options, the program exited without a warning. Fixed.
9. The "Add" button for the 'Add Word' option was accidentally deleted from the Word List Preferences menu. This meant that words could only be added by hitting the return button. Fixed.

Changes:

10. The default file type for Open Dir has been changed to .txt
11. The 'Directory' displays in both the main window and keywords list preferences window have been deleted. As directories and sub-directories can now be used, this feature has become redundant and perhaps confusing.

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:

1. Binding to plot canvas lines to allow jumping to hit in file (same as in concordance view).
2. Binding to file list allowing the user to simply click on the file name when using the File View tool to view the file
3. History feature for all entries (use up and down arrows on keyboard)
4. Much faster concordancer processing (especially sorting... up to 10 times faster)
5. Ability to clear content of tools (Clear Tool, Clear All Tools, Clear All Tools and Files)

6. Ability to save list of all loaded files in settings file
7. 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."
8. New Lemma tool allowing all lemmas of a word list word to be displayed. Note: A lemma list is required.
9. Added bindings to Wordlist Range list to allow words to be deleted (<Delete> button) or kept (<Insert> button) as with other tools.
10. Swap button to switch the main and reference corpora when doing keyword analyses. (Accessible from the Keyword preferences Menu).
11. Fixed serious bug that caused the program to freeze or crash when an incorrectly formed regular expression was used as a search term.
12. Fixed serious bug that caused some word, keyword and font encodings to not be loaded correctly from the user settings files
13. Fixed bug that caused the 'Reset' buttons to not work correctly.
14. 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.
5. Labels in the software (in particular those related to language encodings) have been slightly altered to make them clearer.
6. Improved the warnings given when the Collocates Tool is not in sync with the Word List Tool, during the calculation of collocate significance.
7. 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.
8. 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
5. case setting.
6. 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!).

7. 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.
8. *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.
9. 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 shifts to be the best option.
10. Some users have said they could not open the font selection windows. This problem has been fixed.
11. Many new ordering options have been added, for example, from word ends.
12. 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.
13. Some of the tool names have been changed. In particular "lexical bundles" are now called "N-Grams", which is a more familiar term.
14. 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.
15. 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.
16. This read me file has been largely rewritten. I hope there are fewer spelling mistakes this time!
17. 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 *AntConc*3.0 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 an 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 been re-designed.

2.4.1

New since *AntConc* 2.4.1 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

1. First, progress indicators were added to 'pages' of *AntConc*.
2. Second, a new file view feature was added to view target files in their original state.
3. Third, a keyword generation feature has been added using log-likelihood and chi-squared methods.
4. Finally several bugs were found, in particular, bugs centered on 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

1. 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.
2. Numerous bugs centered on the way the software entered a 'Busy' mode were corrected.
3. The main core of the software was also updated resulting in a quicker, 'cleaner' processing of the data.
4. Performance improvements should be noticed as a result.

2.2.3

1. Updated file and directory selection dialog boxes to run smoothly in a Windows environment.

2. Also, changed the default colors for sort highlighting, and search window frame size.
3. A number of small bugs were also corrected

2.2.2

1. Corrected critical fault with compiler that 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

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

1. 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

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

2.0

1. Added new sort feature, for rearranging concordance lines.
2. 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

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

1.0

First version

Copyright 2021 Laurence Anthony. All rights reserved.