

Readme file for AntWordProfiler 1.3.1
Developed by Laurence Anthony
Center for English Language Education in Science and Engineering
Faculty of Science and Engineering
Waseda University
June 30, 2012

AntWordProfiler is a freeware word profiler. It is written in Perl 5.14 using ActiveState's excellent Komodo and PDK development tools. The program can be launched by simply double clicking on the executable file, which can be downloaded from the Laurence Anthony Laboratory web site: <http://www.antlab.sci.waseda.ac.jp/>. The program can run under any windows environment including Win 98/Me/2000/NT and XP, and also Macintosh OSX and Linux computers. If a user finds any problem launching the program under a particular OS, please let me know.

SIMPLE GUIDE TO GETTING STARTED

Follow the steps below to analyze one or more target texts.

1) Launch AntWordProfiler by double clicking on the icon.

2) Import one or more level lists (using the Choose button)

The level lists are a vertical list of the baseform (and possible families) of the words you wish to investigate. The format of the level lists follows Paul Nation's Range program.

Example:

A
 AN
ABLE
 ABILITY
 ABLER
 ABLEST
 ABLY
 ABILITIES
 UNABLE
 INABILITY

...

NOTE: The baseword will also be included as a family member. In other words, the list above, will be treated as equivalent to the list below.

A
 A
 AN
ABLE
 ABLE
 ABILITY
 ABLER
 ABLEST
 ABLY
 ABILITIES
 UNABLE
 INABILITY

...

Three baseword lists are included with the program. These are:

- 1_gsl_1st_1000.txt
- 2_gsl_2nd_1000.txt
- 3awl_570.txt

3) Import one or more target user files (using the Choose button)

4) Select one or more of the output data options

- a) Statistics: this will produce a set of statistics about the target files.
- b) Word Types: this will produce a list of word types in each target file that are also in the level lists
- c) Word Groups (Families): this will produce a list of base word groups (families) in each target file that are also in the level lists
- d) Include a complete frequency list of all words in the target file together with some valuable statistics
- e) Include words not in list(s): this will include words not in the level list(s) as part of the results
- d) Include words not in file(s): this will include words not in the target file(s) as part of the results

5) Decide on a sort setting:

- a) Sort by word spelling
- b) Sort by word range (document frequency)
- c) Sort by word frequency

6) Decide to process all files together (Batch Process = No) or all files separately (Batch Process = Yes)

7) Press START to generate the results:

8) Select a save option from the FILE menu, to save the results in Paul Nation's Range program format, or tabbed-spaced text format

#####

#####

USING THE USER FILE VIEW

Follow the steps below to view (and simplify) a target file.

- 1) Select a target file from the list of User File(s)
- 2) Click on the 'View' button
- 3) Click the "Level List Words" button to color-highlight the words included in the level lists.
- 4) Click the "Non-Level List Words" button to color-highlight words not included in the level lists.
- 5) Use the 'Prev' and 'Next' buttons to navigate to words that you want to edit or simplify.
- 6) If the <Thesaurus> option is activated in Menu->Settings->Thesaurus Settings, click the 'Edit' button to view synonyms of selected words at different levels. Words that have a thesaurus entry will be underlined. Use the 'Replace' or 'Replace All' buttons to edit the target text.
- 7) Choose one of the 'Save Options' to decide how edited target texts will be saved. The 'None' will leave the text unchanged. The 'Add Level Index Tags' will tag all words with an underbar

followed by a level number.

8) Click the 'Save' button to save the edited target text.

Notes: When using the File View tool, the following features are available:

Click [Return] to choose the selected thesaurus entry in the thesaurus viewer.

Hover over the Level labels in the 'Level Coverage' display to see the full pathname.

#####

KEYBOARD SHORTCUTS

In many windows, the following shortcuts will work:

[ALT .] (alt period): makes the text larger

[ALT ,] (alt comma): makes the text smaller

In the target file VIEW window , the following shortcuts will work:

[ALT n] moves the cursor to the 'n'ext color-highlighted word

[ALT p] moves the cursor to the 'p'revious color-highlighted word

[ALT l]: activates the "Highlight Level List Words" button

[ALT h]: activates the "Highlight Non-Level List Words" button

[ALT u]: makes the non-highlighted words appear underlined

[ALT b]: makes the highlighted words appear in black

[ALT g]: makes the highlighted words appear in grey

[ALT w]: makes the highlighted words appear in white

Other standard shortcut keys, such as CTRL-X (cut), CTRL-C (copy),

CTRL-V (paste), CTRL-Z (undo), CTRL-Y (redo), should work as expected.

#####

ACKNOWLEDGMENTS

I would like to thank Paul Nation of Victoria University of Wellington for being a constant source of excellent suggestions and feedback during the development of version 3.3.x. I would also like to thank him for giving me permission to include his baseword lists in the software. Without these, the value of the software would be severely diminished. In addition, I would like to thank Chris Sheppard of Waseda University for encouraging me to develop this software at the outset, and his suggestions and comments during the early testing stages. The development of AntWordProfiler has been supported by a Grant-in-aid for Scientific Research by the Japan Society for the Promotion of Education, Science, Sports and Culture, Japan (No. 18700658). Development of version 3.3 has also been helped by generous support from Compass Media, Seoul.

LEGAL MATTER

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

KNOWN ISSUES

None.

REVISION HISTORY

1.3.1

This is a minor update addressing one small bug and adding one or two new features

New Features

1) Added an option in the thesaurus settings menu to allow users to select to automatically update the file view after

editing it with the thesaurus tool.

2) Massively improved (>10 times) the speed of the file viewer tool when highlighting texts.

Bug fixes

- 1) Corrected a bug that caused the local frequency count in the file viewer to report values for characters as well as words.
- 2) Tidied the interface a little to large frequency values to become hidden in the file viewer tool
- 3) Fixed a bug that caused the program to hang if the Profiler Tool start button was clicked when no files or lists were imported.

1.3.0

This is a major update adding new features and correcting various bugs in the 1.2x versions.

New Features

- 1) Revised the names and cleaned contents of the internal baseword lists (approved by Paul Nation)
- 2) Added an option to view a complete frequency list of the target files in the profile
- 3) Added an option to view words not in the target files in the profile
- 4) Added new sort options (word, range, freq)
- 5) Added a batch process option (main window only)
- 6) Revised this readme file
- 7) Added a new thesaurus viewer (activated via a menu option)
- 8) Revised the profile output to include cumulative frequencies
- 9) Improved the User File Viewer tool to display all level lists in a scrolled pane.
- 10) Added new editing functions (Previous, Next) to the User File Viewer tool
- 11) Added a new edit function to simplify the target text via a level-marked thesaurus.

Bug fixes

- 1) Corrected an unintuitive feature that duplicated the listing of internal baseword lists after importing a settings file.
- 2) Corrected a bug that caused the windows to not be centered when first displayed.
- 3) Corrected a bug that prevented more than eight baseword lists to function correctly.

1.2.1

This is a minor update to correct a few bugs and minor issues in the 1.200 version.

New Features

- 1) Added a new icon to the software.

Bug fixes

- 1) Fixed an issue that caused the global settings options to not appear on Macintosh OS X.

Minor issues

- 1) The icon for AntWordProfiler has been update to a graphic that matches other AntLab tools. I hope you like it!
- 2) The height of widgets has now been fixed to make them fluid and thus allow them to take their full height on each operating system.

1.200

This version includes many changes and so deserves its 1.2 status. A list of new features and changes are below.

New Features

- 1) Redesigned the software engine so that the program now works smoothly with Unicode (UTF-8) encoded files. This means the program will work with any language (including Asian languages) provided that the target files, and baseword lists are saved in the UTF-8 encoding.
- 2) Redesigned the settings files to be easier to edit and work with (for creating user generated settings files). The settings file is no longer an XML file. Rather, it is a specil file with a very simple format.
- 3) The main window title now shows if the program is using a user settings or not.
- 4) Any errors do not show appear on the main window anymore. This makes debugging a little more difficult, but improves the user experience.

1.104

Bug fixes

1) Fixed code to correctly update the level list word database when analyzing new texts. Previous versions would produce strange results if the User File view was used without generating a main set of results first.

1.103

New Features

1) Recoded to allow Macintosh OSX version to be compiled easily.

Bug fixes

None

1.102

Bug fixes

1) Deleted two file menu options that were redundant (Choose Main Word List, Choose User Word List)

2) Added widget names to interface that were included in the settings file but not reflected.

3) Repositioned widgets to allow non-English names to be used successfully.

4) Revised the internal level lists to exactly match those in Paul Nation's original Range program (not Range BNC). The previous versions of AntWordProfiler used the first three Range BNC lists. Note that either set of lists can be imported and used as is. Note also that the placement of some entries in the Range lists is questionable (see Bug fix 1 in version 1.101).

1.101

Bug fixes

1) Corrected the placement of "M" in the Nation Baseword List 1. Originally this had it's own entry, but now it is correctly placed as a family member of "BE"

2) Corrected the trimming of non-token characters from the beginning of file lines. These caused the not-in-list results to include a stray entry count.

3) Updated and redefined the settings of the compiler for AntWordProfiler. This may resolve the problems reported by some users in Asia who cannot launch the program.

4) Added XML header to user settings file

1.100

New Features

1) Introduced a database back engine, allowing much larger files to be processed.

2) Hugely improved the speed of processing. Now, 7 million words can be processed in around 1 minute.

3) Introduced a tag ignore feature allowing tagged files to be processed without change.

4) Greatly simplified the program (and interface) by adopting Paul Nation's Range format for level list files.

5) Importing and exporting of the settings is now possible via an XML file.

6) Simplified the save format options. The Excel save format may be reintroduced later.

7) Changed the name of the program to more accurately represent what it does.

8) Re-wrote this readme file to reflect the new changes.

Bug fixes

1) Several mistakes in the language of the interface have been corrected.

2) Errors reports that occur when pressing Cancel in some dialog boxes have been fixed.

1.01

Bug fixes

1) Corrected how to read in main word lists and level lists so that blank lines and lines starting with non-token characters are ignored. In version 1.00, blank lines would generate very strange results.

2) Added character encodings for Chinese, Taiwanese, and Korean (Japanese encodings were already included in version 1.00)

1.00 This is the first version of the program.

Copyright Laurence Anthony 2012