

Readme file for AntWordProfiler1.200 (Previously named AntVocabCheck)
Developed by Laurence Anthony
Center for English Language Education in Science and Engineering
Faculty of Science and Engineering
Waseda University
June 27th, 2009

AntWordProfiler is a freeware word profiler. It is written in Perl 5.10 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

...

(The first three Nation baseword lists are included with the program)

- 3) Import one or more target user files (using the Choose button)
- 4) Select one or more of the output data options
 - a) Word Types: this will produce a list of word types in each target file that are also in the level lists
 - b) Word Groups (Families): this will produce a list of base word groups (families) in each target file that are also in the level lists

- c) Include words not in list(s): this will include words not in the level lists as part of the results
- d) Statistics: this will produce a set of statistics about the target files.

6) Press START to generate the results:

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

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 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

In the target file VIEW window , the following shortcuts will work when highlighting non-level list words:

- [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 Chris Sheppard of Waseda University for encouraging me to develop this software, and his suggestions and comments during the testing stages. I would also like to thank Paul Nation for giving me permission to include his baseword list in the software. Without this, the value of the software would be severely diminished. 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).

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

- 1) Due to a bug in Microsoft Windows, it seems that opening files directly on the "desktop," can cause the program to suddenly crash. To be safe, I strongly recommend users to run the program under a pathname with no aliases, and only containing ASCII characters.
- 2) Some users in Asia have reported problems launching the program on their system. The problem seems to be related to dll file conflicts. I am now investigating the problem. This version attempts to address the problem by updating the compiler and adding more dll files to the executable.
- 3) To prevent the internally stored level lists from multiplying after importing a user settings file, disable the automatic loading of the internal files via the Global Settings-File option.

REVISION HISTORY

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 special 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 its 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 2008