

Readme file for AntVocabCheck1.01  
Developed by Laurence Anthony  
Center for English Language Education in Science and Engineering  
Waseda University  
September 12th, 2007

AntVocabCheck is a freeware vocabulary profiler. It is written in Perl 5.8 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 AntVocabCheck by double clicking on the icon.
- 2) Import a baseword list to covert inflections to their baseword form. On launching the software, you will find that Paul Nation's baseword list is already included. (<http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx>). If you want to use your own baseword list, just follow the format of the baseword list included.

- 3) Import one or more level lists (using the Choose button)  
The level lists are a vertical list of the baseform of the words you wish to investigate. Lists such as JACET 1000-8000 can be used directly.

Example:

the  
a  
of  
go  
come  
study  
...

- 4) Import one or more target user files (using the Choose button)
- 5) Select one or more of the output data options
  - a) words in level lists: this will produce a list of base words in each target file that are also in the level lists
  - b) words in level lists (cum): this will produce a list of base words in each target file that are also in the current level list and all previous level lists, in other words, a cumulative result. So, the results for level list 1 will be identical to a) but for level 2, the words in BOTH level 1 and level 2 will be outputted. This allows you to easily see which words are in the target file UP TO a certain level.
  - c) Lists minus words: this will produce a list of base words that are in the level lists but NOT in the target files
  - d) Lists minus words (cum): this will produce a cumulative list of words that are in the level lists but NOT in the target files (see b) above for more details)
  - e) Words not in lists (cum): this will produce a list of base words that are in the target files but NOT in the level lists. This list is effectively the inverse of list b) above.
  - f) Word Stats: this will produce a set of statistics about the target files.

- 6) Press START to generate the results:

5a) If the target files contain words that are not in the basewords list, the program will ask you to save them as a user

baseword file. Follow the instructions in the program to do this. Then, hit START again.

7) Select a save option from the FILE menu, to save the results in raw, tabbed-spaced text, csv, or Excel .xls 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 AntVocabCheck is 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

AntVocabCheck 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

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.

## REVISION HISTORY

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 2007