



ProtAnt (Windows, MacOS, Linux)

Build 1.3.1

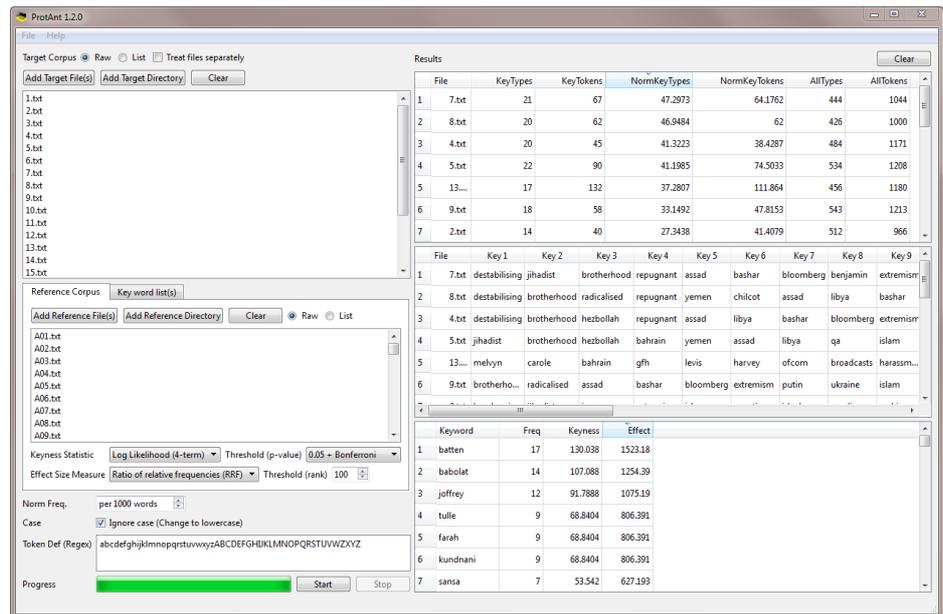
Laurence Anthony, Ph.D.

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.

August 2, 2023

Introduction

ProtAnt is a freeware prototypical text detection tool developed in collaboration with Paul Baker of Lancaster University, UK. *ProtAnt* takes a corpus of texts (**UTF-8 encoded**) and compares them either individually or as a whole against a reference corpus (**UTF-8 encoded**) or list of 'key' words (**UTF-8 encoded**) to find characteristic features in the target files. Then, *ProtAnt* looks at each individual target file and counts how many of these characteristic features are in them. The target files are then ranked in terms of their prototypicality by the number of characteristic features they contain. *ProtAnt* runs on any computer running Microsoft Windows (tested on Windows 10), Apple MacOS (Intel/Silicon) (tested on OSX 10.15 Catalina and OSX Ventura), or Linux (tested on Linux Mint 17). *ProtAnt* is developed in *Python* and *Qt* using the *PyInstaller* compiler to generate executables for the different operating systems.



Getting Started

Windows - Installer

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

Macintosh OS X

Double click the *ProtAnt.dmg* file to create a ProtAnt disk image on your desktop. Open the disk image and drag and drop the ProtAnt 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 *ProtAnt.tar.gz* file into a folder of your choice. In the *ProtAnt* folder, double click the *ProtAnt.sh* file to launch the software. On the command line, type `./ProtAnt.sh` to launch the software.

Finding prototypical texts using a reference corpus

Step 1: Select the type of "Target Corpus" files you are going to use (raw files or word lists)

Step 2: Select if you want keywords to be generated by comparing all target file together with the reference files ("Treat files separately" - unchecked option) or each target file separately with the reference files ("Treat files separately" - checked option)

Step 3: Select the target corpus files that you want to analyze. You can do this in three ways:

- 1) Click on the File->Open Target File(s) menu option or the "Add Target File(s)" button below the "Target Corpus" label and select the files you want to analyze;
- 2) Click on the File->Open Target Dir menu option or the "Add Target Directory" button below the "Target Corpus" label and select a directory of files you want to analyze;
- 3) Drag and drop files directly onto the *ProtAnt* application.

Step 4: Select the reference corpus files that you want to analyze. You can do this in three ways:

- 1) Click on the File->Open Reference File(s) menu option or the "Add Reference File(s)" button below the "Reference Corpus" label and select the files you want to add;
- 2) Click on the File->Open Reference Dir menu option or the "Add Corpus Directory" button below the "Reference Corpus" label and select a directory of files you want to add;
- 3) Drag and drop files directly onto the *ProtAnt* application.

Step 5: Choose the keyness statistic, keyness threshold value (p-value), effect size measure, and effect size threshold (the cutoff rank) you would like to use to generate keywords for the target corpus. Note that a keyness threshold of 0 means all (positive) keywords will be considered, and an effect size threshold of -1 means that all effect values will be shown. A positive keyword means that it is relatively more frequent in the target corpus than the reference corpus. Negative keywords (relatively more frequent in the reference corpus than the target corpus) are not considered.

Step 6: Choose the normalization function for frequencies displayed in the results window.

Step 7: Choose to ignore case in the target corpus (change all words to lowercase)

Step 8: Decide the ordering of keywords (in the bottom right pane), which will also determine the ordering of keywords shown per file in the middle pane.

Step 10: Decide a suitable token definition based on regular expression (regex) syntax.

Step 9: Click "Start" to begin the analysis.

Finding prototypical texts using 'key' word lists

Step 1: Select the type of target corpus files you are going to use (raw files or word lists)

Step 2: Select the target corpus files that you want to analyze. You can do this in three ways (see Step 3 above):

Step 3: Select the 'key' word list files that you want to analyze. You can do this in three ways:

- 1) Click on the File->Open Keywords File(s) menu option or the "Add Keywords File(s)" button below the "Key word list(s)" label and select the files you want to add;
- 2) Click on the File->Open Keywords Dir menu option or the "Add Keywords Directory" button below the "Key word list(s)" label and select a directory of files you want to add;
- 3) Drag and drop files directly onto the *ProtAnt* application.

Step 4: Choose the normalization function for frequencies displayed in the results window.

Step 5: Choose to ignore case in the target corpus (change all words to lowercase)

Step 6: Decide a suitable token definition based on regular expression (regex) syntax.

Step 7: Click "Start" to begin the analysis.

Note 1: If you click on the File->Close Target Files menu option, the File->Close Reference Files menu option, or the File->Close Keywords Files menu option, the files will be removed from the relevant list.

Note 2: If you click on the "Clear" button below the "Target Corpus" label, the "Key word list(s)" label, or the "Clear" button below the "Reference Corpus" label, or the "Clear" button below the "Key word list(s)" label, the files will be removed from the relevant list.

Note 3: The results of the prototypical text detection are shown in the top right window. The keywords appearing in each (ranked) corpus file are shown in the middle right window. The complete list of keywords is shown in the bottom right window. All columns can be sorted in either ascending or descending order by clicking on the column headers.

Note 4: The prototypical text detection analysis can be stopped at any time by clicking the "Stop" button.

Note 5: If you are using target corpus or reference corpus word lists, they should be formatted as RANK, FREQ, TYPE separated by tabs. Any line beginning with # will be ignored.

Note 6: If you are using 'key' word files, they should be formatted with each 'key' word on a new line.

Additional Features

The following shortcuts are available as is standard on the operating system:

Windows

CTRL-A ⇒ Select All

CTRL-C ⇒ Copy

CTRL-V ⇒ Paste

Macintosh

Command-A ⇒ Select All

Command-C ⇒ Copy

Command-V ⇒ Paste

CITING/REFERENCING *ProtAnt*

Use the following method to cite/reference *ProtAnt* according to the APA style guide:

Anthony, L. and Baker, P. (YEAR OF RELEASE). *ProtAnt* (Version VERSION NUMBER) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/>

For example if you download *ProtAnt 1.2.0*, which was released in 2016, you would cite/reference it as follows:
Anthony, L. and Baker, P. (2016). *ProtAnt* (Version 1.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/>

Note that the APA instructions are not entirely clear about citing software, and it is debatable whether or not the "Available from ..." statement is needed. See here for more details:

<http://owl.english.purdue.edu/owl/resource/560/10/>

KNOWN ISSUES

None at present.

Copyright 2023 Laurence Anthony. All rights reserved.