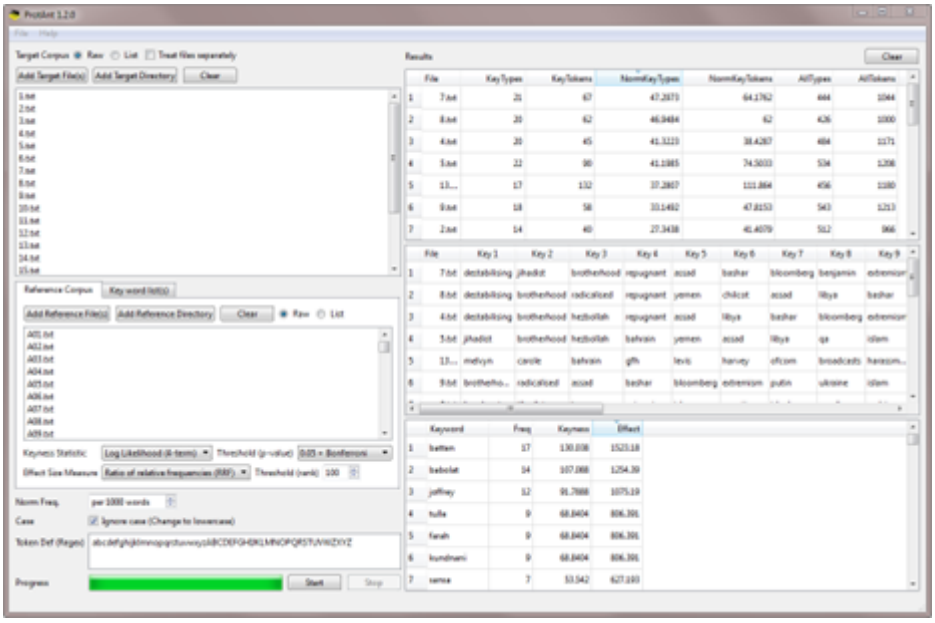
ProtAnt

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

Table of Contents

Introduction



ProtAnt is a freeware prototypical text detection tool developed in collaboration with Paul Baker of Lancaster University, UK. *ProtAnt* compares a target corpus against either a reference corpus or one or more keyword lists, identifies the characteristic features of the target data, and then ranks target files by how strongly those features appear in each text.

ProtAnt can work with UTF-8 plain-text files and list-formatted input files. It runs on Microsoft Windows, macOS, and Linux, and is developed in Python with Qt.

Getting Started

Windows (Installer)

Double-click the *ProtAnt installer package* and follow the installer instructions. After installation, you can launch *ProtAnt* from the Start menu.

Windows (Portable)

Unzip the *ProtAnt portable ZIP package* into a folder of your choice. In the *ProtAnt* folder, double-click the *ProtAnt.exe* file to launch the program.

macOS

Double-click the *ProtAnt disk image* to mount it. Open the disk image and drag the *ProtAnt* app to the Applications folder (or another location if you prefer). You can then launch the app from Applications or Launchpad.

Linux

Download the *ProtAnt Flatpak bundle* and install it with `flatpak install <bundle-file>`. After installation, you can launch *ProtAnt* from your desktop application menu or from the command line with `flatpak run org.antlab.ProtAnt`.

Finding Prototypical Texts Using a Reference Corpus

1. Select the type of *Target Corpus* files you want to use:
    - raw files
    - word lists
2. Decide whether keywords should be generated:
    - from all target files together compared with the reference corpus
    - from each target file separately compared with the reference corpus
3. Load the target corpus files in one of three ways:
    - use `File -> Open Target File(s)...`
    - use `File -> Open Target Dir...`
    - drag and drop files onto the application
4. Load the reference corpus files in one of three ways:
    - use `File -> Open Reference File(s)...`
    - use `File -> Open Reference Dir...`
    - drag and drop files onto the application
5. Choose:
    - the keyness statistic
    - the keyness threshold value
    - the effect size measure
    - the effect size cutoff rank
6. Choose the normalization value for frequencies displayed in the results window.
7. Choose whether to ignore case in the target corpus.
8. Decide the ordering of the keyword table in the bottom-right pane. This ordering also determines the keyword order shown per file in the middle pane.
9. Decide on a suitable token definition using regular expression syntax.
10. Click **Start** to begin the analysis.

Useful Hints

- A keyness threshold of `0` means all positive keywords are considered.
- An effect size threshold of `-1` means all effect values are shown.
- Positive keywords are items that are relatively more frequent in the target corpus than in the reference corpus.
- Negative keywords are not included in the results.

Finding Prototypical Texts Using Keyword Lists

1. Select the type of target corpus files you want to use:
    - raw files
    - word lists
2. Load the target corpus files in one of the standard ways described above.
3. Load the keyword list files in one of three ways:
    - use `File -> Open Keywords File(s)...`
    - use `File -> Open Keywords Dir...`
    - drag and drop files onto the application
4. Choose the normalization value for frequencies displayed in the results window.
5. Choose whether to ignore case in the target corpus.
6. Decide on a suitable token definition using regular expression syntax.
7. Click **Start** to begin the analysis.

Input Formats

Corpus Word Lists

If you are using target corpus or reference corpus word lists, each row should be tab-separated in this format:

`RANK<TAB>FREQ<TAB>TYPE`

Lines beginning with # are ignored.

Keyword Files

If you are using keyword files, place one keyword per line.

Results

The results area contains three tables:

- the ranked target files
- the keywords found in each ranked file
- the full keyword list with frequency, keyness, and effect values

All result columns can be sorted in ascending or descending order by clicking the column headers.

Managing Files

- If you use `File -> Close Target Files`, `File -> Close Reference Files`, or `File -> Close Keyword List`, the corresponding files are removed from the list.
- If you click a **Clear** button below one of the file-list sections, the files in that section are removed from the list.
- Analysis can be stopped at any time by clicking the **Stop** button.

Shortcuts

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

Windows

Shortcut	Action
Ctrl+A	Select All
Ctrl+C	Copy
Ctrl+V	Paste

macOS

Shortcut	Action
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., & Baker, P. (YEAR OF RELEASE). *ProtAnt* (Version VERSION NUMBER) [Computer software]. Waseda University. http://www.laurenceanthony.net/

For example, if you download *ProtAnt* 1.4.0, which was released in 2026, you would cite/reference it as follows:

Anthony, L., & Baker, P. (2026). *ProtAnt* (Version 1.4.0) [Computer software]. Waseda University. http://www.laurenceanthony.net/

Known Issues

None at present.