



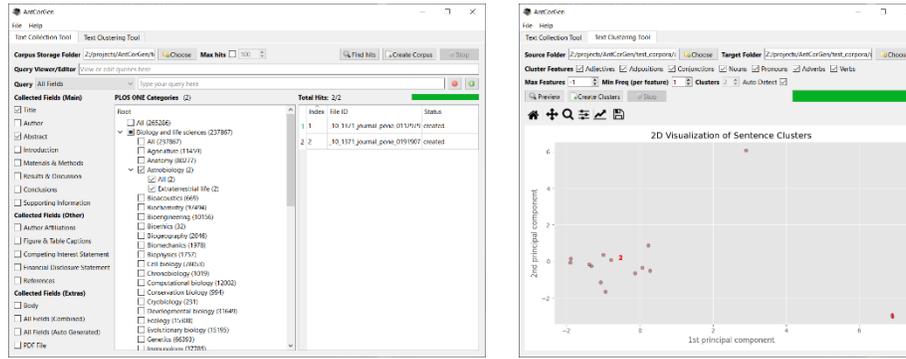
# AntCorGen (Windows, MacOS, Linux)

## 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

January 10, 2026



## Introduction

*AntCorGen* is a freeware corpus generation tool. *AntCorGen* lets you search for documents in the PLOS ONE research database via search queries and/or subject category browsing and decide which sections (e.g. title, abstract, introduction) of these documents should be stored. *AntCorGen* then accesses the database, downloads the sections, and saves each one as a text file in an appropriate folder. *AntCorGen* can also analyze the different parts of speech (e.g. adjectives, verbs) of words in the files and cluster similar sentences into sub-groups. These sub-groups will show similar patterns of language use.

*AntCorGen* runs on any computer running Microsoft Windows (tested on Win 7), Macintosh OS X (tested on OS X 10.9 Mavericks), and Linux (tested on Linux Mint 17) computers. It is developed in Python and Qt using the *PyInstaller* compiler to generate executables for the different operating systems.

## Getting Started (No installation necessary)

### Windows - Installer

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

### Macintosh OS X

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



**Step 5:** Click the "Preview" button to show a scatter-plot visualization of the clusters. If the clusters are not separated in the visualization, adjust the features and parameters as necessary. The scatter-plot can be resized, zoomed, label-adjusted, and saved using the icons above the plot image.

**Step 6:** Click the "Create Clusters" button to cluster the document sentences and store them in the target folder.

**Step 7:** Click the "Stop" button to stop the clustering process at any time.

## NOTES

### Comments/Suggestions/Bug Fixes

All new editions and bug fixes are listed in the revision history below. However, if you find a bug in the program, or have any suggestions for improving the program, please let me know and I will try to address the issues in a future version.

This software is available as 'freeware' (see Legal Matter below), but it is important for my funding to hear about any successes that people have with the software. Therefore, if you find the software useful, please send me an e-mail briefly describing how it is being used.

### CITING/REFERENCING *AntCorGen*

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

Anthony, L. (YEAR OF RELEASE). *AntCorGen* (Version VERSION NUMBER) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/>

For example if you download *AntCorGen 0.0.1* which was released in 2017, you would cite/reference it as follows: Anthony, L. (2017). *AntCorGen* (Version 1.0.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/>

### KNOWN ISSUES

None at present.

Copyright: Laurence Anthony 2024