



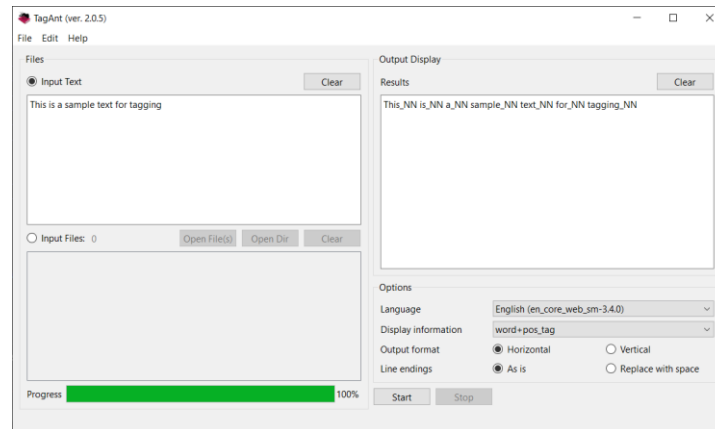
# TagAnt (Windows, MacOS, Linux)

## Build 2.1.1

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.

March 9, 2024



## Introduction

*TagAnt* is a freeware, multi-language tagging tool built on top of the *SpaCy* natural language processing (NLP) framework. It takes either direct input (typed into the interface) or an input list TEXT files (.txt) (**UTF-8 encoded**), Microsoft WORD (.docx) files, and/or PDF (.pdf) files and processes them to generate the following components for each word: WORD, PART-OF-SPEECH (POS) NAME PART-OF-SPEECH (POS) TAG, LEMMA. Options allow the choice of language, the component information to be displayed, the output format, and the handling of line-endings. *TagAnt* runs on any computer running Microsoft Windows (built on Win 10), MacOS (built on Mac Catalina), and Linux (built on Linux Mint). *TagAnt* 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 *TagAnt.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 *TagAnt.zip* file into a folder of your choice. In the *TagAnt* folder, double click the *TagAnt.exe* file to launch the program.

### Macintosh OS X

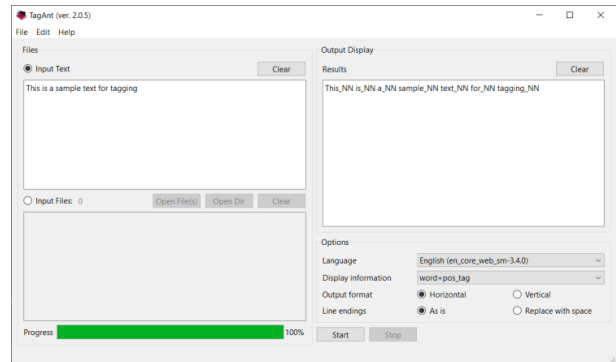
Double click the *TagAnt.dmg* file to create a *TagAnt* disk image on your desktop. Open the disk image and drag and drop the *TagAnt* app onto the Applications folder (or into another location if you desire). You can then launch the app by double clicking on the icon in the Applications folder or the Launchpad.

### Linux

Decompress the *TagAnt.tar.gz* file into a folder of your choice. In the *TagAnt* folder, double click the *TagAnt.sh* file to launch the software. On the command line, type `./TagAnt.sh` to launch the software.

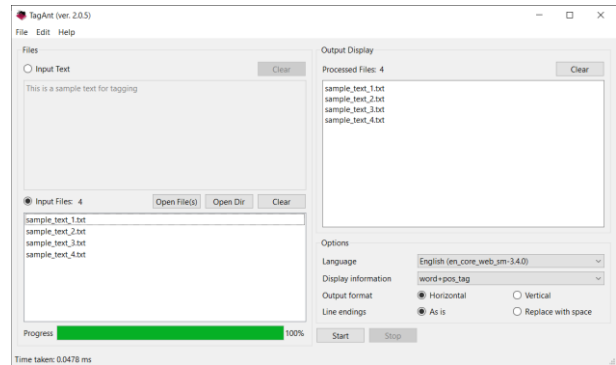
## Tagging Input Text

- Step 1:** Under "Files", select the "Input Text" radiobox on the left of the main window.
- Step 2:** Enter the text that you want to process.
- Step 3:** Under "Options" choose the settings to use (see the Options section below for further details)
- Step 4:** Click "Start" to begin the tagging process. The tagging process can be stopped at any time by clicking the "Stop" button.



## Tagging Input Files

- Step 1:** Under "Files", select the "Input Files" radiobox on the left of the main window.
- Step 2:** From the "File" menu, select "Open File(s)" or "Open Dir". Choose the files you want to process in the Dialog box. You can load TEXT files (.txt) (**UTF-8 encoded**), Microsoft WORD (.docx) files, and PDF (.pdf) files.
- Step 3:** Under "Options" choose the settings to use:
- Step 4:** Click "Start" to begin the tagging process. The tagging process can be stopped at any time by clicking the "Stop" button.



## Options

The following options are available during processing.

Language:

Select the language to be processed. The following languages are available:

Bokmål, Catalan, Chinese, Croatian, Danish, Dutch, English, Finnish, French, German, Greek, Italian, Japanese, Korean, Lithuanian, Macedonian, Polish, Portuguese, Romanian, Russian, Slovenian, Spanish, Swedish, Ukrainian, Universal.

Display information:

Select the word components to be shown. Combinations of the following are possible:

WORD, POS name, pos TAG, LEMMA

Output format

Select the output format to be shown. The "Horizontal" (Default) option will show the output matching the input on a line-by-line basis. In this case, the word components are combined with underscores (e.g. dog\_NN). The "Vertical" option will show the components for every word of the input on separate lines. In this case, the word components are separated by tabbed spaces.

Line endings:

Select how line endings should be processed. The "As is" (Default) option leaves the line endings as is, with no changes. The "Replace with space" option replaces all line endings with spaces, rendering the complete file as a single line file.

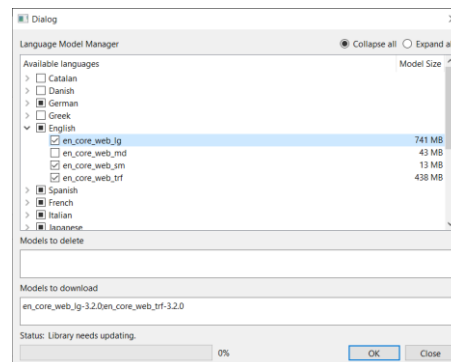
## Additional Features

The output display can be selected and copied via the Edit menu. The following shortcuts are also available as is standard on the operating system:

Windows:	CTRL-A ⇨ Select All	CTRL-C ⇨ Copy	CTRL-V ⇨ Paste
Macintosh:	CMD-A ⇨ Select All	CMD -C ⇨ Copy	CMD -V ⇨ Paste

## Tagging Input Text

TagAnt comes with several pre-installed *SpaCy* language models, including those for English, French, German, Italian, Japanese, and Chinese. If you want to tag your data using other language models, you can use the Language Model Manager to install them. Access the Language Model Manager via the File Menu. In the language model viewer, select language models to add/remove and then click OK to update the language model library.



## CITING/REFERENCING *TagAnt*

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

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

For example, if you download *TagAnt 2.0.0*, which was released in 2021, you cite/reference it as follows:  
Anthony, L. (2021). *TagAnt* (Version 2.0.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/>

## KNOWN ISSUES

- 1) If a very large file (e.g., greater than your computer's available memory) is copied to the "Input Text" box, it can cause the software to crash. Similarly, once your file is tagged and the results are generated, displaying the results and/or copying them might again use up all the available system memory and cause the program to crash. To avoid this problem, very large files should always be tagged using the "Input File(s)" option, so that they are not displayed while being processed.