

Nalaprop 1.0b9 for Mojave
Release Notes
Howard Oakley <https://eclecticlight.co>

These release notes describe the features of **Nalaprop**, an app which uses macOS 10.14 Mojave's natural language processing to parse text, mark it up according to detected parts of speech ('lexical classes'), generate classified frequency lists of words used, and more.

One of many new features in Mojave, which were detailed and demonstrated at WWDC 2018, built-in support for parsing natural language is one of the most exciting. Primarily intended to support natural (spoken) language interaction with iOS devices, its tools will be of great value to many macOS users interested in language.

Nalaprop is an exploration of natural language parsing which brings to your Mac what was demonstrated at WWDC.

What you need

- A Mac running Mojave. This release doesn't run on any earlier releases of macOS, which lack the features which it relies on.
- A copy of the latest release of Nalaprop from <https://eclecticlight.co/downloads/> (This is delivered by secure HTTPS download.)

Getting started

Nalaprop comes compressed as a Zip file, which you should decompress, and move the app to your preferred folder, such as /Applications. It is not fussy where it is run from, though.

Nalaprop not only uses my developer signature, but is notarized for Mojave. It cannot access documents in private areas of your Mac's storage, such as emails or Photos libraries. If you experience any difficulties opening the app the first time, please contact me immediately.

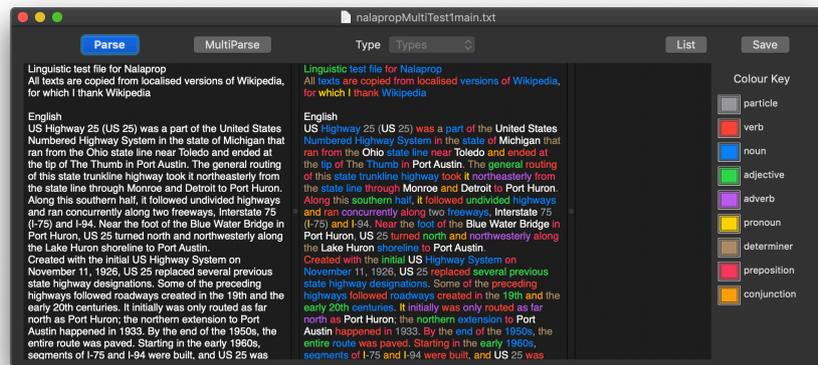
This is the seventh beta release of Nalaprop, and has changed significantly since the previous release.

Basic use

To open a plain text or rich text (RTF) file, use the **Open...** command in the **File** menu. To paste in plain text, if necessary create a new window using the **New** command, then paste that text into the lefthand pane of any document window. Rich text files are read in as plain text, as Nalaprop will impose its own styling to their text.

When you open an existing document, it will automatically be parsed into the centre pane, but word lists will not be generated at that stage. To re-parse the contents of the lefthand pane at any time, click on the **Parse** button or press Command-Return ⌘↵. Currently, parsing is performed in the main thread, which means that large documents such as a Charles Dickens novel can take some time to parse, and the spinning beachball cursor may appear for that period. A future version will perform parsing in the background to alleviate this issue.

Nalaprop 1.0b9 for Mojave
Release Notes
Howard Oakley <https://eclecticlight.co>



Nalaprop then uses Mojave to parse the text and display the marked-up text in the centre.

Interactive editing

This makes it easy to use Nalaprop interactively: adjust the window contents so that the left and centre views occupy as much of the width as possible, and type your text into the lefthand view, pressing  whenever you want to see it parsed. Unlike the centre and righthand views, you can change the font and size in the lefthand view to make editing most comfortable for you.

MultiParse

To examine your text in more detail, click on the **MultiParse** button and Nalaprop will parse it several times over to extract fuller details about each word within it. Again, this is currently performed in the main thread, and for longer texts the spinning beachball cursor may appear while parsing.



When you have MultiParsed the text, the **Type** popup menu is enabled, and offers five options for display of the centre pane:

- **Script** - this colours each word according to the script detected, e.g. Latn for Latin;
- **Language** - this colours each word according to the language detected, e.g. ru for Russian;
- **Types** - this colours each word according to part of speech using an arbitrary scheme derived from the types of word present;
- **Types (std)** - this colours each word by part of speech according to Nalaprop's standard scheme, which is intended to aid analysis;

Nalaprop 1.0b9 for Mojave

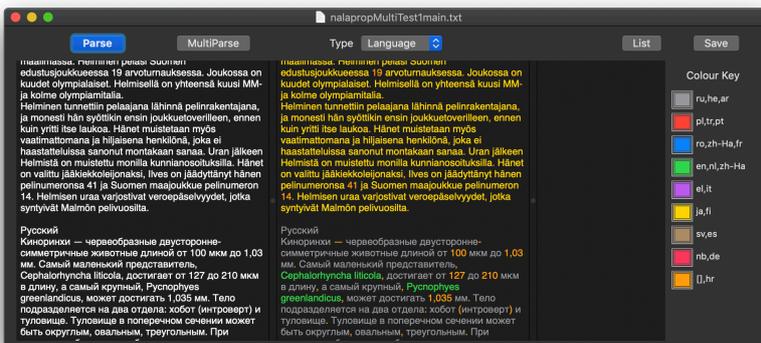
Release Notes

Howard Oakley <https://eclecticlight.co>

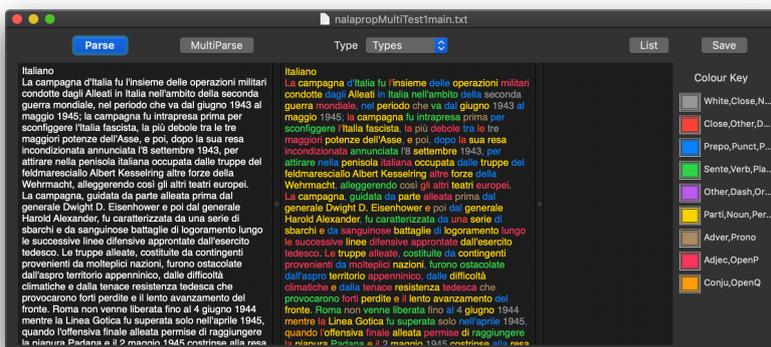
- **Lemma** - this gives a stem or base form of those words which it can, e.g. in French, *le*, *la* and *les* are given the lemma *le*. Those stems are then coloured using the same scheme as **Types**. This option loses punctuation and text layout but can be valuable when reading a language which you know less well.



The **Script** setting distinguishes each underlying script, here Latin and Arabic.

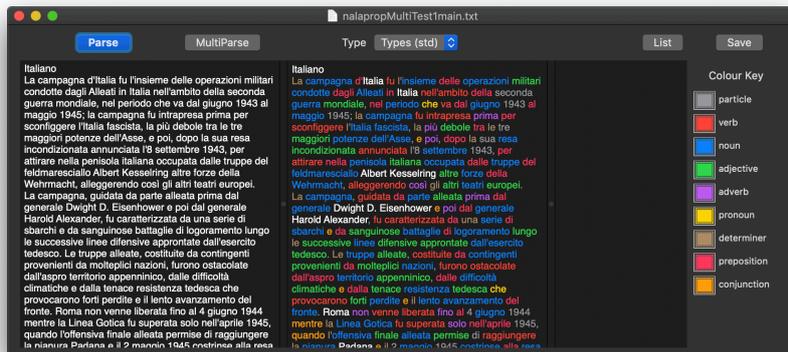


The **Language** setting distinguishes each language. Note that numerals are given a type of [], indicating that they are independent of the underlying language.

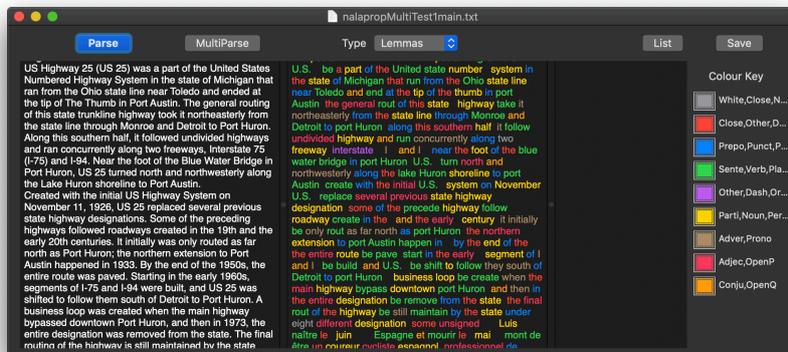


The **Types** setting colours each word according to its lexical type (part of speech), using an arbitrary colour scheme. The colour key currently doesn't assign all the types detected.

Nalaprop 1.0b9 for Mojave
Release Notes
Howard Oakley <https://eclecticlight.co>



The **Types (std)** setting colours each word according to its lexical type, using an optimised colour scheme intended to make grammatical structure clear.



The **Lemmas** setting shows the 'root' forms of words which it knows, coloured according to the Types scheme.

Colour keys

The **Colour Key** shown in the main window gives an abbreviated list of the types corresponding to the colours, which should be sufficient for examining the script and language. In most samples of text there are more than nine word types which are detected. To make it easier to see the full colour key, use the **Custom colour key** command in the **Help** menu. This opens an accessory window showing in full all the current colour assignments for the window in which you last clicked the **MultiParse** button.



Nalaprop 1.0b9 for Mojave
Release Notes
Howard Oakley <https://eclecticlight.co>

Because you can have many of these palette windows open at any time, the title of each matches that of its parent window. The key given within it remains frozen from the time that the window is opened, and shows the key for the current selected **Type** in that window. This may seem complicated, but gives you great flexibility in the colour keys that you have on display.

Nalaprop re-uses existing colours in a circular fashion for those additional types: the tenth word type, for example, uses the first colour again, and the eleventh the second colour. If you want to examine the lexical classes or parts of speech, you should find the **Types (std)** menu option preferable, as it allocates colours on an optimised and pre-determined basis. However, it doesn't offer as many classes as the other **Types** option.

The standard colour scheme for showing parts of speech (**Types (std)** in the menu, and that shown when you first open a document or click on the Parse button) uses the following colour mappings:

- nouns, blue
- verbs, red
- adjectives, green
- adverbs, purple
- pronouns, yellow
- determiners, brown
- particles, grey
- numbers, grey
- prepositions, pink
- conjunctions, orange
- classifiers, orange
- idioms, green
- interjections, orange
- personal, organisational, and place names, white or black (according to Light/Dark mode at the time)
- others, including punctuation and any unparsed text, grey.

You can view this colour key in full detail using the **Standard colour key** command in the **Help** menu. This opens a fixed-size window showing the colours and their significance.



Nalaprop can fully parse English, French, Spanish, German, Italian, Portuguese, Russian, Turkish, and additional languages as supported by Mojave. Many other languages are supported in part, and it can recognise most scripts in general use, and many languages. Support for lexical class analysis is currently limited to those eight languages, and for lemmas in even less. Mojave is expected to see an increasing range and quality of support over the coming months.

Find

Nalaprop now supports multiple **Find** features which effectively turn it into an interactive concordance. Because each of the three text views has quite different contents, Find works in two slightly different ways according to which is active.

To search the text in the editable view at the left, click on the text within it and use the **Find** command in the **Edit** menu. This adds a standard **Find** bar to the top of that view, into which you can type the text which you want to locate.

To search the parsed text in the centre view, which you cannot edit, click in that view and use the **Find** command in the **Edit** menu. This opens the standard **Find** window with its full set of tools.

To search the wordlists in the righthand view, which you also cannot edit, click in that view and use the **Find** command in the **Edit** menu. This adds a standard **Find** bar to the top of that view, into which you can type the text which you want to locate.

If you already have a **Find** window open, you can click in any of the three text views to perform that search within that view. You can't, though, use either **Find** bar to search either of the other text views.

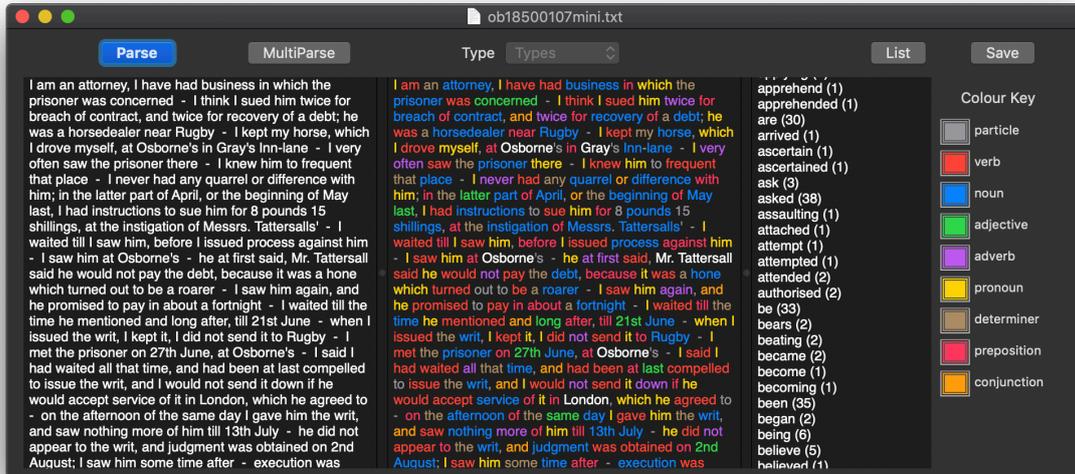
To use Nalaprop as a concordance, select a word from the wordlists at the right, copy and paste it into the **Find** window. Then click in the centre view, and find each of its occurrences. If you want to search for lemmatised forms, remember to switch the **Types** menu to **Lemmas** first.

Word Frequency Counts

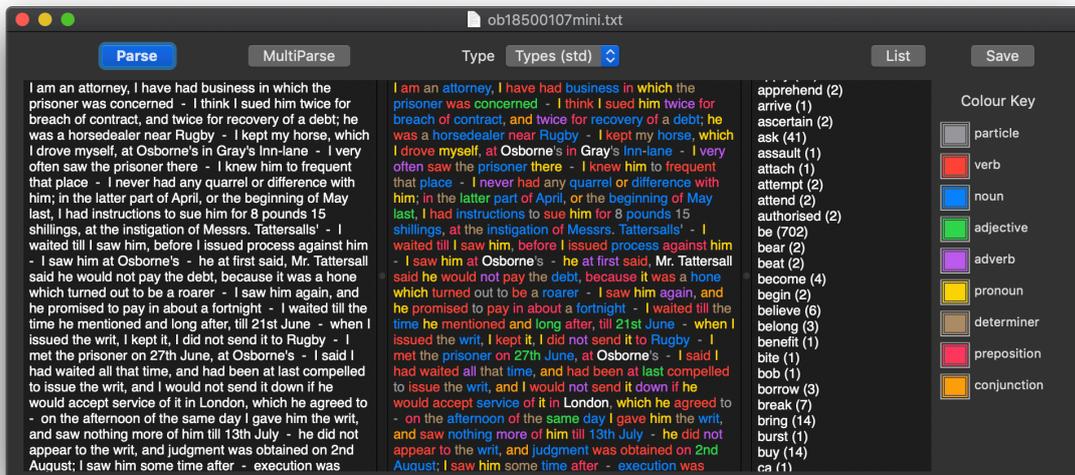
To perform word frequency analysis on the parsed text, click on the **List** button at the right. This runs through the parsed text, and displays the numbers of word occurrences according to the lexical class of the word.

Two types of frequency analysis are available: when you have just clicked on the **Parse** button, or have just read in a document, simple analysis is performed on the words found in the text. In this, each occurrence of *is*, for example, is counted as the verb form *is*.

Nalaprop 1.0b9 for Mojave
Release Notes
Howard Oakley <https://eclecticlight.co>



When you have just clicked on the **MultiParse** button and the **Type** menu is enabled, the analysis is performed on lemmatised versions of each word, where they are available. So each occurrence of *is* will be counted as an occurrence of the lemma *be*. Where a lemmatised form is not available for a word, the word itself is used instead.



To switch between the two forms of analysis, use the **Parse** (for regular) and **MultiParse** (for lemmatised) buttons before clicking on **List**.

Word frequency analysis shown at the right is performed as follows:

- every word parsed into one of the major lexical classes is sorted into its class
- with the exception of names (which are left with case unchanged), those words are reduced to lower case
- an alphabetical listing of words is then given, with the number of occurrences of that word in the text, for each lexical class in turn (in the case of lemmatised forms, the lemma is substituted for the word, when available)

Nalaprop 1.0b9 for Mojave
Release Notes
Howard Oakley <https://eclecticlight.co>

- at the end of each class listing, the total number of words of that class found in the the text is given
- at the end of all the class listings, the total number of words assigned to those classes is given.

At the moment, words from different languages which are used in the text are pooled together.

The **Save...** command has been disabled to help prevent you from overwriting original files. When you **Save as...**, select the Rich text (RTF) option to preserve colour coding in the parsed output. Saving only saves the text displayed in the centre panel, not the original text in the lefthand one. To save the frequency analysis in the righthand panel, use the **Save** button above it.

To print the text in any of the three panels, click on the text in that panel so that the insertion cursor is within it, then use the **Print** command.

The text file `nalapropMultiTest1.txt` contains sample passages of twenty different languages, copied from Wikipedia's localised versions, for which I am very grateful to Wikipedia.

Technical information

Swift 4.2 code used to achieve the parsing includes:

```
import NaturalLanguage

let theMASOut = NSMutableAttributedString()
let tagger = NLTagger(tagSchemes: [.nameTypeOrLexicalClass])
tagger.string = theStr
let options: NLTagger.Options = []
tagger.enumerateTags(in: theStr.startIndex..
```

Why the name Nalaprop? It's a pun on Mrs Malaprop, and **Natural Language Processing**.

Nalaprop 1.0b9 for Mojave
Release Notes
Howard Oakley <https://eclecticlight.co>

Change list

1.0b9:

- added support for Find in each of the three text views
- added palette view for colour keys to script, language, lexical class
- miscellaneous interface tweaks.

1.0b8:

- added Cmd-Enter shortcut to Parse button to make parse-while-edit simple.

1.0b7 (unreleased):

- reorganised button logic and added List button
- ported to Swift 4.2.1 in Xcode 10.1
- improved labels on colour swatches
- added lemmatised wordlist.

1.0b6:

- major structural changes: nows opens with parsed Types view
- Parse an opened document now only adds wordlists
- MultiParse to support multiple parsed views, e.g. Script, Language
- added in-window colour key
- added internal multi-tagged version of text to support these.

1.0b5:

- fixed typos in the word frequency analyses.

1.0b4 (unreleased):

- added third scrolling text view
- added classified word frequency analysis to righthand view, and Save button.

1.0b3:

- tested against 20 different language passages, from Arabic to Turkish (included)
- added Colour Key help window.

1.0b2:

- added support for opening RTF files as plain text
- enabled printing support
- built using Xcode 10.0 release.

1.0b1:

- tidied the interface, removing superfluous controls
- notarized for Mojave
- built using Xcode 10β5.

Nalaprop 1.0b9 for Mojave
Release Notes
Howard Oakley <https://eclecticlight.co>

1.0a2:

- fixed a crash when language not recognised/supported yet
- added second stage parsing with popup menu.

1.0a1:

- initial release
- crashes when trying to parse Chinese.

4 November 2018.