

Amanuensis: Generic V1.0

(last review: 13 November 2020)¹

1 Purpose

The Amanuensis (www.riedlberger/amanuensis) story started with an application for searching in a more or less complete database of Latin juristic texts. The first program, then called Amanuensis (now renamed as “Amanuensis: Roman Law”) combines this huge database with a nifty search interface.

The search interface proved so useful that it was recycled for the database of the Latin portions of the Acts of the Ecumenical councils (this is “Amanuensis: ACO”). In other words, “Amanuensis: Roman Law” and “Amanuensis: ACO” are more or less identical, excepting for the underlying database.

The present program completes the Amanuensis family of applications. “Amanuensis: Generic” enables you to search through *any* text (as long as you have it as plain text file) using the powerful Amanuensis search engine.

Maybe you wonder: why should you use Amanuensis to search through a plain text file, while you simply could press CTRL+F?

First, Amanuensis offers advanced search options, such as “*” (truncating), “+” (find an additional word which is not directly adjacent but nearby), “-” (exclude text blocks with a certain word) or “?” (one unspecified character).

Second, Amanuensis presents search results in a clearly arranged manner. Instead of jumping through a document by repeatedly performing your search, you have a one-stop list of results. A double click takes you immediately to the context (and from there back again).

Third, Amanuensis has specific adaptations for Latin and Greek. In Latin texts, assimilation is automatically taken care of (*summ** finds *submovere*, for instance), in Greek texts, you automatically type Greek letters in the search box, and Amanuensis ignores any diacritics while searching.

Typical use cases of “Amanuensis: Generic” are searching through texts you have scanned yourself; or which you found somewhere in the depths of the internet; or for which the existing search routines are not comfortable (or quick) enough.

2 License

Amanuensis itself is freeware. You can use it for free (even in commercial contexts). You can also pass it on for free. However, you must in this case always clearly state the authors’ names. And remember, there is no warranty or liability whatsoever. The license is “Creative Commons Attribution-NoDerivatives 4.0 International Public License.”

Note that the Windows Amanuensis version includes some open source code with different licenses, namely:

- The Code Project Open License 1.02 (www.codeproject.com/info/cpol10.aspx)
 - *A Pretty Good Splash Screen in C#* by Tom Clement

¹ This piece of software is part of a project that has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program under grant agreement No. 677638.

(www.codeproject.com/Articles/5454/A-Pretty-Good-Splash-Screen-in-C)

- MIT License
(<https://github.com/grammophone/Grammophone.BetaImport/blob/master/LICENSE.md>)
 - *Grammophone.BetaImport* by grammophone
(<https://github.com/grammophone/Grammophone.BetaImport>)
- GNU Lesser General Public License (<http://opensource.org/licenses/lgpl-3.0.html>)
 - *Fast Colored TextBox for Syntax Highlighting* by Pavel Torgashov
(www.codeproject.com/Articles/161871/Fast-Colored-TextBox-for-syntax-highlighting)
 - *gettext-cs-utils* by Manas Technology Solutions
(<http://code.google.com/p/gettext-cs-utils>)

Further, Windows Amanuensis ships with two free monospaced typefaces (polytonic Greek looks simply awful when displayed using Courier New):

- DejaVu Sans Mono (<http://sourceforge.net/projects/dejavu>)
- Ubuntu Monospace (<http://font.ubuntu.com>)

3 Using Amanuensis: Generic on modern texts

Using the application is best explained by a hands-on exercise. As example, download the text of the index volume of A. H. M. Jones's *Later Roman Empire* in plain text version. You find at https://archive.org/stream/JonesLaterRomanEmpire03/Jones_Later-Roman-Empire03_djvu.txt (if you type that in your browser, save it by selecting File, Save As, and save it as TXT).

In Amanuensis, click the "About" tab. Leave the radio button at "Unspecified," as Jones writes in English, which obviously is neither Latin nor Greek. Click "Load Text" and select your downloaded text.

In the "Search" tab, you can now start your searches. The main rules are:

- Two words divided by a blank: you only get hits with exactly that sequence (so "but is" finds attestations only of *that sequence*, 8 instances in our example text)
- "+" operator: the second word must be near-by but not necessarily adjacent (so "but +is" finds 194 instances)
- the "-" operator excludes the specified word (e.g., "but +is -still" has 182 instances)
- you can truncate words by using "*" so "Symm*" finds both "Symmachus" and "Symm."
- comma, period, semicolon, colon and colon are ignored; so "Symm" finds "Symm."

If you would like to see concrete examples for these operators, please refer to the detailed instructions of "Amanuensis: Roman Law." There, you will also find an explanation of the two advanced operators "?" and "refining minus" which most users will probably never need.

Why did I choose the index volume of LRE as example? There is no *index locorum* for LRE, which is most unfortunate. There is, for example, no full commentary for the *Codex Theodosianus* later than Godefroy's 17th c. one. While Jones explains "somewhere" most or all CTh. texts with some relation to administration or public law in general, it can be (given the lack of an *index locorum*) a daunting task to find these explanations. Enter "Amanuensis: Generic". Imagine you seek information on anti-heretical constitutions which are contained in the title CTh. 16.5. You could search for "CTh xvi v" (remember, Amanuensis ignores full-stops and commas), but it is better to search for "CTh +xvi +v" because, if Jones lists several CTh texts, he does not repeat "CTh," so it might be separated from

“xvi.” In this way, you can see within a few minutes if and what Jones wrote about a given CTh fragment. [A caveat is in order: this particular scan is bad, just as many other ones you will find on the internet. For example, “VII” is regularly mis-OCR’ed as “vn”; so if you are interested in CTh book VII, you will have to search for both “CTh +vii” and “CTh +vn”.]

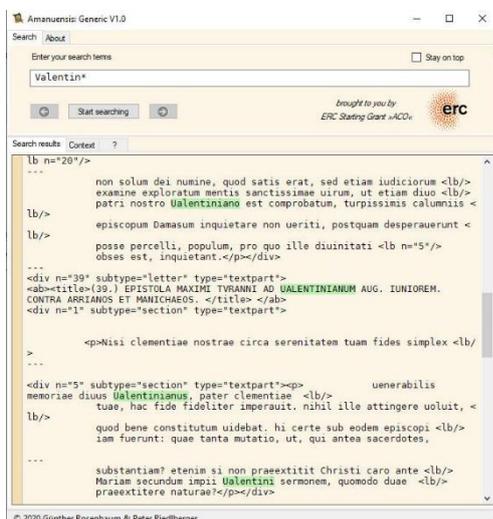
4 Using Amanuensis: Generic on Latin texts

Possibly you wonder by now: why do you need to choose “Unspecified,” “Latin,” or “Greek” when loading a text? Couldn’t be a Latin text searched through just as the Jones LRE example? Indeed, it could, but if you load a Latin text as “Latin,” you get some additional advantages.

Again, let’s try that with a real-life example. If you are interested in the Collectio Avellana, you probably know that this text is missing from the standard databases of Brepolis. But there is one carefully scanned version available at: <https://raw.githubusercontent.com/OpenGreekAndLatin/csel-dev/master/data/stoa0045a/stoa001/stoa0045a.stoa001.opp-lat1.xml> (sorry, I don’t know a more convenient like). Go there and save the text as (e.g.) “Avellana.txt”. Now click the “About” tab of Amanuensis: Generic, select “Latin” and click “Load Text.” This dialog windows pops up:

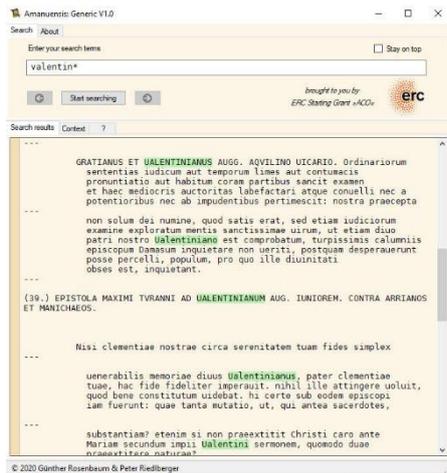


Try “No” / “Nein”, search (e.g.) for *Valentin** and see what happens.



You can see that there are countless tags (that’s the bits within <>), declaring the structure of the text. This is immensely useful if you wish to display the contents in a formatted way; but for our intents, the multitude of tags is rather distracting.² So, let’s go back to the “About” tab, let us “Load Text” again, but this time, click “Yes” / “Ja” and remove all tags.

² If you want to know more about these encoding techniques and their benefits, check e.g. <https://en.wikipedia.org/wiki/EpiDoc>.



Looks much better, doesn't it? Note a few things, however:

- (i) If you remove the tags, this has no effect on the file you load, i.e., it will not be changed. Removing tags only changes the way it is displayed in Amanuensis: Generic. In other ways: you cannot "damage" your file by using this option.
- (ii) Be careful: sometimes, text contained within <...> isn't a tag, but an editorial substitution for a perceived lacuna. If you remove *anything* between <...>, such substitutions will also be gone.
- (iii) In the screenshot above, you can see "letter 39." In this particular version of the *collectio Avellana*, the letter numbers are *not* contained within the tags, so they "survive" the tag removal. This means that even after removal, you can identify the location of any hit by double clicking it to see the context and scrolling a little bit up or down in the context until you see a letter number. *But there are other encoding styles!* Sometimes, the numbering scheme is only to be found *within tags*, so after removal, you have no way to ascertain the position of your result. In these cases, you must leave the tags. If they are too bothersome, you might do your rough searching without tags, and when you have your results (e.g., on a piece of paper), you could load again with tags.

However, choosing Latin as input language has more effects than giving you the option of removing tags. Specifically, choosing Latin entails that:

- no distinction between V/U and v/u (*Valentin** finds *Ualentianus*; *uulgus* finds *vulgus*)
- appended *-que* / *-ve* is ignored
- consonant assimilation is automatically taken care of (e.g., *subm-* finds *summ-* and conversely; for more information, cf. the instructions to Amanuensis: Roman Law)
-

5 Using Amanuensis: Generic on Greek texts

For Greek there is TLG, one might be inclined to say: TLG is practically complete, and there is hardly anything which cannot be accomplished with its search routine. Indeed, that's true—but many users find it very inconvenient to use, and there is no doubt that it is slow. In case you do not to search through the whole corpus of ancient literature but only through a given text, you might prefer Amanuensis: Generic.

First, of course, you need the text. Download `hopper-texts-GreekRoman.tar.gz` which you can find at various places in the internet, e.g. <https://ancientgreekocr.org/archived/hopper-texts-GreekRoman.tar.gz>. This is most of ancient Latin and Greek literature in one gargantuan archive. Don't worry, this is completely legal, as their authors have been dead for quite a while. Unpack this file (on Windows you will need third party software, such as, e.g. Total Commander). In the resulting folder "Classics," you will have all your ancient authors sorted alphabetically.

If you open any of the Greek text files, you will see that they are not encoded in Greek polytonic Unicode, but rather in betacode. So the first sentence of Arrian's *Anabasis* reads not:

Πτολεμαῖος ὁ Λάγου καὶ Ἀριστόβουλος ὁ Ἀριστοβούλου ὅσα μὲν ταῦτ᾽ ἄμφω περὶ Ἀλεξάνδρου τοῦ Φιλίππου συνέγραψαν, ταῦτα ἐγὼ ὡς πάντῃ ἀληθῆ ἀναγράφω, ὅσα δὲ οὐ ταῦτά, τούτων τὰ πιστότερα ἐμοὶ φαινόμενα καὶ ἅμα ἀξιαφηγητότερα ἐπιλεξάμενος.

but rather:

```
*ptolemai=os o( *la/gou kai\ *)aristo/boulos o( *)aristobou/lou o(/sa me\n tau)ta\ a)/mfw peri\
*)aleca/ndrou tou= *fili/ppou sune/grayan, tau=ta e)gw\ w(s pa/nth| a)lhqh= a)nagra/fw, o(/sa de\
ou) tau)ta/, tou/twn ta\ pisto/tera e)moi\ faino/mena kai\ a(/ma a)ciafhgho/tera e)pileca/menos.
```

Fear not! For if you load text as specify it as "Greek," you have two options: first, you can (just like in Latin) remove tags; second, you can "Decode all text to polytonic Greek," i.e., convert from Betacode to polytonic Greek.

So, let's do this! Click the "About" tab, select "Greek," and "Load text." Click "No" / "Nein" regarding tag removal³ but "Yes" / "Ja" regarding "Decode all text to polytonic Greek." Navigate to our file (Classics, Arrian, opensource, `arrian.anab_gk.xml`), and load it.

If you go back to the search tab, anything is now ready for searching in Greek:

- any keyboard input is displayed with Greek letters (using the well-known Betacode mapping)
- diacritics are ignored (thus, α finds any "α", no matter which accents or spirits go with it)

What if you do not want to search through Arrian's *Anabasis* alone, but through all of his works? Simple: Copy/paste the other Arrian files into one gigantic text file, save it, and load it into Amanuensis: Generic. Amanuensis: Generic is well equipped to handle quite large text files, so you can really put your "own" research corpus together (in this way, the database of Amanuensis: Roman Law came into existence).

6 Scanning texts yourself

This section is a small parenthesis, and not directly related to "Amanuensis: Generic." If you wish to use the program with downloaded texts, you may skip this.

Scanning texts is not that difficult as it may seem if you never tried to do it. Any standard scanner will do. Actually, the scanning itself is not the problem, the OCR ("optical character recognition") is. Your software should know how various letter shapes look like, and it should be capable of rectifying minor "reading" mistakes by comparing the result to an in-built dictionary. Luckily, you are not the first person with these needs, and there is powerful software available on the internet:

<https://ancientgreekocr.org/>

<https://dctthree.github.io/antigraphheus/>

³ This particular file has very few texts, and, more importantly, they hold the only clues regarding the numbering schema. So, if you removed them, you would not be able to know the location of a result you found.

7 More features

Amanuensis: Generic has inherited a multitude of features from the original Amanuensis (now called: Amanuensis: Roman Law), such as the “Stay on top” feature or the search history. I am too lazy to list them here. Most or all are self-explanatory, and all of them are explained in the instructions to “Amanuensis: Roman Law.”

8 Final remarks

If you detect any errors or if you have ideas for additional functionality that might be implemented rather painlessly (be reasonable, Amanuensis is freeware after all!), please contact us. Furthermore, if you like the program and use it, you might as well just drop me an e-mail. A happy Amanuensis user somewhere on the globe who sends me a nice e-mail might just make my day! The contact address is peter@riedlberger.de.

9 Version history

13 November 2020

V1.0

- First release version