



1. Installing NooJ for Windows

A step-by-step guide

This series of tutorials is based upon work from COST Action
Multi3Generation CA18231, supported by COST
(European Cooperation in Science and Technology).

COST (European Cooperation in Science and Technology) is a funding agency for research and innovation networks. Our Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career and innovation, cf. www.cost.eu

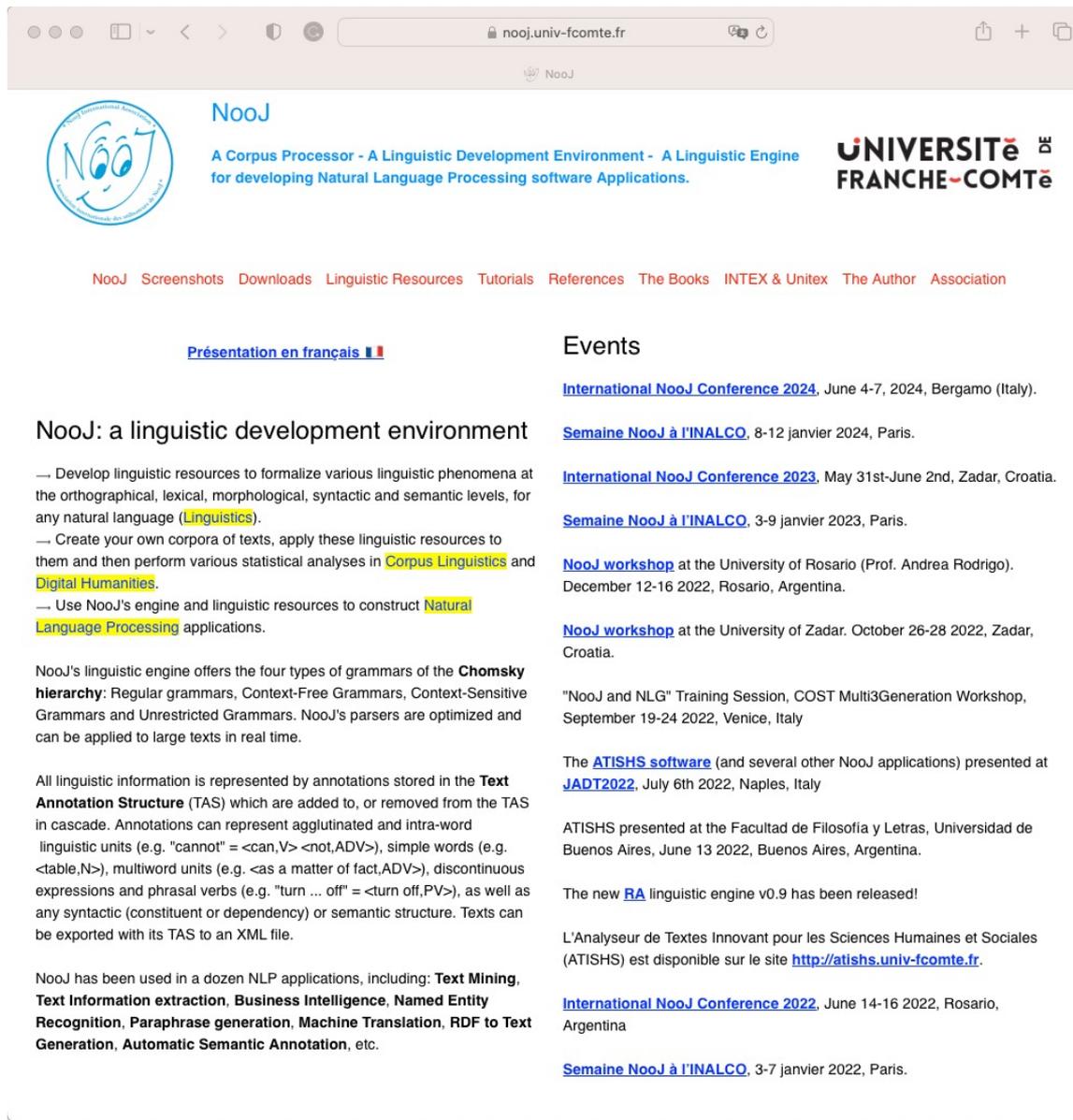
Installing NooJ

The four variants

1. **NooJ for Windows: the complete toolbox. Runs on Windows (.NET technology).**
2. NooJ for Java: a simplified version of NooJ for demos and students of Software Engineering: well-documented open-source (JAVA technology)
3. ATISHS for macOS: a French application in the Digital Humanities, various statistical tools (macOS technology)
4. RA for Linux, macOS, and Windows: the new engine, used in the WEB application <https://webnooj.univ-fcomte.fr> (Swift, Vapor technology)

NooJ for Windows: Step 1

Go to <https://nooj.univ-fcomte.fr>



The screenshot shows a web browser window displaying the NooJ website. The browser's address bar shows the URL nooj.univ-fcomte.fr. The website header features the NooJ logo on the left, which is a circular emblem with a stylized face and the text 'Linguistic Development Environment' and 'Université de Franche-Comté'. To the right of the logo is the text 'NooJ' and a descriptive sentence: 'A Corpus Processor - A Linguistic Development Environment - A Linguistic Engine for developing Natural Language Processing software Applications.' Further right is the logo for 'UNIVERSITÉ DE FRANCHE-COMTÉ'. Below the header is a navigation menu with links: 'NooJ', 'Screenshots', 'Downloads', 'Linguistic Resources', 'Tutorials', 'References', 'The Books', 'INTEX & Unitex', 'The Author', and 'Association'. The main content area is divided into two columns. The left column has a section titled 'Présentation en français' with a small flag icon, followed by the heading 'NooJ: a linguistic development environment'. Below this heading are three bullet points describing the environment's capabilities. The right column has a heading 'Events' followed by a list of conferences and workshops with their dates and locations. At the bottom, there is a paragraph about the linguistic engine's capabilities and a list of applications where NooJ has been used.

[Présentation en français](#)

NooJ: a linguistic development environment

- Develop linguistic resources to formalize various linguistic phenomena at the orthographical, lexical, morphological, syntactic and semantic levels, for any natural language ([Linguistics](#)).
- Create your own corpora of texts, apply these linguistic resources to them and then perform various statistical analyses in [Corpus Linguistics](#) and [Digital Humanities](#).
- Use NooJ's engine and linguistic resources to construct [Natural Language Processing](#) applications.

NooJ's linguistic engine offers the four types of grammars of the **Chomsky hierarchy**: Regular grammars, Context-Free Grammars, Context-Sensitive Grammars and Unrestricted Grammars. NooJ's parsers are optimized and can be applied to large texts in real time.

All linguistic information is represented by annotations stored in the **Text Annotation Structure (TAS)** which are added to, or removed from the TAS in cascade. Annotations can represent agglutinated and intra-word linguistic units (e.g. "cannot" = <can,V> <not,ADV>), simple words (e.g. <table,N>), multiword units (e.g. <as a matter of fact,ADV>), discontinuous expressions and phrasal verbs (e.g. "turn ... off" = <turn off,PV>), as well as any syntactic (constituent or dependency) or semantic structure. Texts can be exported with its TAS to an XML file.

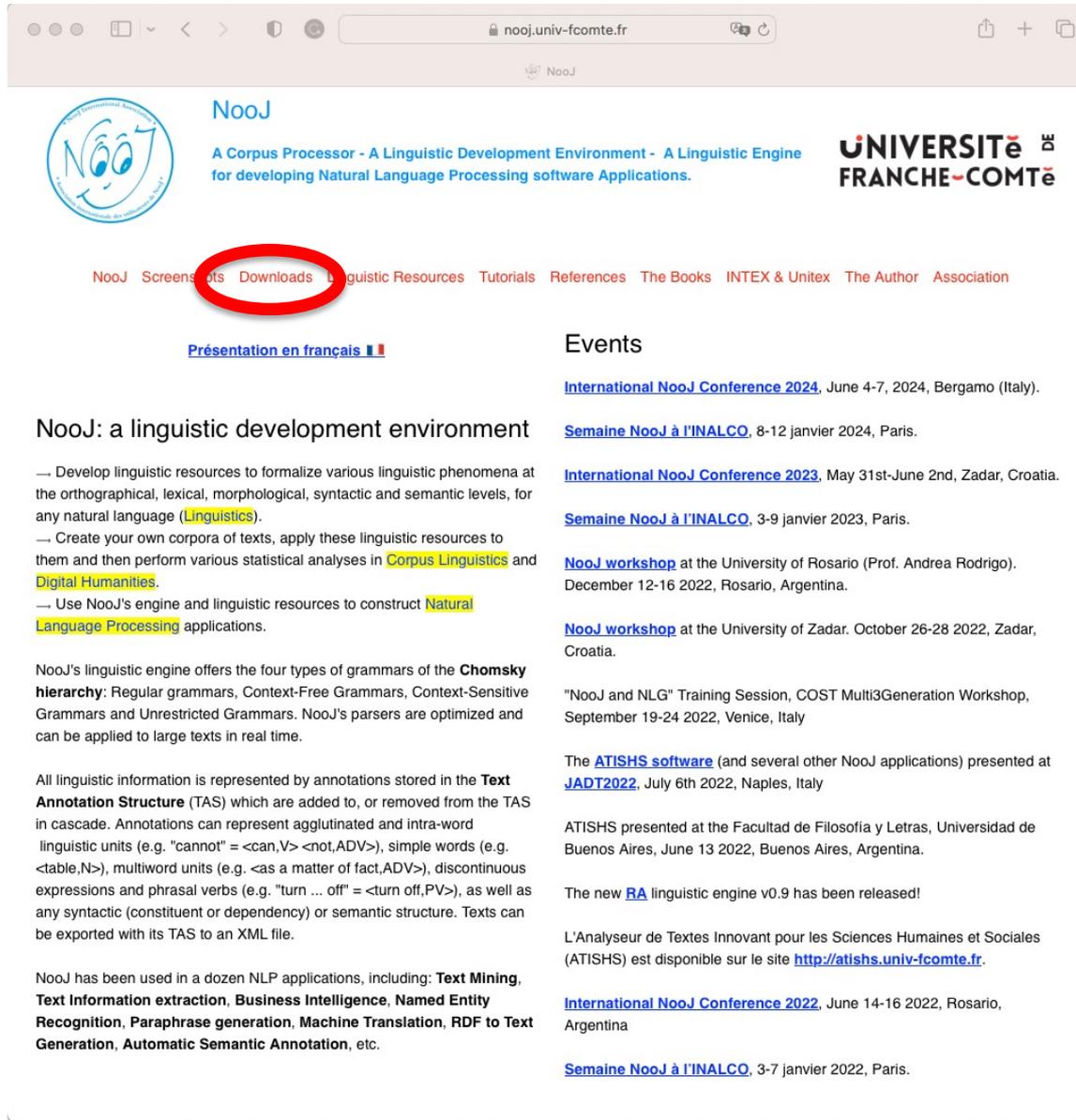
NooJ has been used in a dozen NLP applications, including: **Text Mining**, **Text Information extraction**, **Business Intelligence**, **Named Entity Recognition**, **Paraphrase generation**, **Machine Translation**, **RDF to Text Generation**, **Automatic Semantic Annotation**, etc.

Events

- [International NooJ Conference 2024](#), June 4-7, 2024, Bergamo (Italy).
- [Semaine NooJ à l'INALCO](#), 8-12 janvier 2024, Paris.
- [International NooJ Conference 2023](#), May 31st-June 2nd, Zadar, Croatia.
- [Semaine NooJ à l'INALCO](#), 3-9 janvier 2023, Paris.
- [NooJ workshop](#) at the University of Rosario (Prof. Andrea Rodrigo). December 12-16 2022, Rosario, Argentina.
- [NooJ workshop](#) at the University of Zadar. October 26-28 2022, Zadar, Croatia.
- "NooJ and NLG" Training Session, COST Multi3Generation Workshop, September 19-24 2022, Venice, Italy
- The [ATISHS software](#) (and several other NooJ applications) presented at [JADT2022](#), July 6th 2022, Naples, Italy
- ATISHS presented at the Facultad de Filosofía y Letras, Universidad de Buenos Aires, June 13 2022, Buenos Aires, Argentina.
- The new [RA](#) linguistic engine v0.9 has been released!
- L'Analyseur de Textes Innovant pour les Sciences Humaines et Sociales (ATISHS) est disponible sur le site <http://atishs.univ-fcomte.fr>.
- [International NooJ Conference 2022](#), June 14-16 2022, Rosario, Argentina
- [Semaine NooJ à l'INALCO](#), 3-7 janvier 2022, Paris.

NooJ for Windows: Step 1

Go to <https://nooj.univ-fcomte.fr>



NooJ
A Corpus Processor - A Linguistic Development Environment - A Linguistic Engine
for developing Natural Language Processing software Applications.

UNIVERSITÉ DE
FRANCHE-COMTÉ

[NooJ](#) [Screenshots](#) [Downloads](#) [Linguistic Resources](#) [Tutorials](#) [References](#) [The Books](#) [INTEX & Unix](#) [The Author](#) [Association](#)

[Présentation en français](#)

Events

[International NooJ Conference 2024](#), June 4-7, 2024, Bergamo (Italy).

[Semaine NooJ à l'INALCO](#), 8-12 janvier 2024, Paris.

[International NooJ Conference 2023](#), May 31st-June 2nd, Zadar, Croatia.

[Semaine NooJ à l'INALCO](#), 3-9 janvier 2023, Paris.

[NooJ workshop](#) at the University of Rosario (Prof. Andrea Rodrigo).
December 12-16 2022, Rosario, Argentina.

[NooJ workshop](#) at the University of Zadar. October 26-28 2022, Zadar, Croatia.

"NooJ and NLG" Training Session, COST Multi3Generation Workshop,
September 19-24 2022, Venice, Italy

The [ATISHS software](#) (and several other NooJ applications) presented at
[JADT2022](#), July 6th 2022, Naples, Italy

ATISHS presented at the Facultad de Filosofía y Letras, Universidad de
Buenos Aires, June 13 2022, Buenos Aires, Argentina.

The new [RA](#) linguistic engine v0.9 has been released!

L'Analyseur de Textes Innovant pour les Sciences Humaines et Sociales
(ATISHS) est disponible sur le site <http://atishs.univ-fcomte.fr>.

[International NooJ Conference 2022](#), June 14-16 2022, Rosario,
Argentina

[Semaine NooJ à l'INALCO](#), 3-7 janvier 2022, Paris.

NooJ: a linguistic development environment

— Develop linguistic resources to formalize various linguistic phenomena at the orthographical, lexical, morphological, syntactic and semantic levels, for any natural language ([Linguistics](#)).

— Create your own corpora of texts, apply these linguistic resources to them and then perform various statistical analyses in [Corpus Linguistics](#) and [Digital Humanities](#).

— Use NooJ's engine and linguistic resources to construct [Natural Language Processing](#) applications.

NooJ's linguistic engine offers the four types of grammars of the **Chomsky hierarchy**: Regular grammars, Context-Free Grammars, Context-Sensitive Grammars and Unrestricted Grammars. NooJ's parsers are optimized and can be applied to large texts in real time.

All linguistic information is represented by annotations stored in the **Text Annotation Structure (TAS)** which are added to, or removed from the TAS in cascade. Annotations can represent agglutinated and intra-word linguistic units (e.g. "cannot" = <can,V> <not,ADV>), simple words (e.g. <table,N>), multiword units (e.g. <as a matter of fact,ADV>), discontinuous expressions and phrasal verbs (e.g. "turn ... off" = <turn off,PV>), as well as any syntactic (constituent or dependency) or semantic structure. Texts can be exported with its TAS to an XML file.

NooJ has been used in a dozen NLP applications, including: **Text Mining**, **Text Information extraction**, **Business Intelligence**, **Named Entity Recognition**, **Paraphrase generation**, **Machine Translation**, **RDF to Text Generation**, **Automatic Semantic Annotation**, etc.

NooJ for Windows: Step 2

go to “Downloads” page

Latest version: v7.1 — Latest build: May 2022 —
(The new RA linguistic engine has been presented at the NooJ 2022 conference, see below to download workbench and presentation).

— [Download NooJ for Windows here.](#)

Download the **NooJApp.zip** file, move it onto your desktop, double-click it to open it, and then drag the **NooJApp** folder onto your desktop to uncompress it. Note the compressed file **NooJApp.zip** to make sure you will not use it accidentally. Inside the folder **NooJApp**, inside the folder **_App**, double-click the file **NooJ.exe** (wait a first, be patient to answer "OK" to all questions), and then to launch NooJ.

If you get the message "Windows protected your PC", click "More info", and then "Run anyway".

You might want to add a shortcut for the **NooJ.exe** file to the desktop or dock and associate NooJ file types (.noc, .nof, .nog, .nom, .not) to the NooJ application so that you can launch NooJ by double-clicking NooJ linguistic resources, see **Info** — **Preferences**.

Note to MacOS and Linux users: NooJ runs fine in virtual machines such as VirtualBox, VMWare Fusion and Paralels.

— [Download NooJ for Java here.](#)

The Java NooJ version is a simplified version of NooJ developed with the European Metashare CESAR project's support for pedagogical applications, both for students in computer science (to study the source code) and students in computational linguistics: it is [open-source](#) and very well documented, its file formats are readable in text formats. It runs on any operating system, including Linux, MacOS, Windows and UNIX.

Make sure a contemporary version of JAVA as well as its JDK are correctly installed on your computer.

Uncompress the Java NooJ.zip file onto your desktop, and then delete the .zip file to make sure you will not use it accidentally.
Run a shell (e.g. Terminal on a Mac), then type in the two commands:

```
cd "Desktop/Java NooJ"  
java -jar NooJ.jar
```

Attention Mac users: the first time you launch Java NooJ, you might get warning messages because both the JAVA JDK and NooJ applications were downloaded from the Internet instead of from Apple's AppStore. You might need to bypass the warnings and/or allow the application to run in the Mac's System Preferences — Security and Privacy.

Click here →
to download
file:
NooJApp.zip

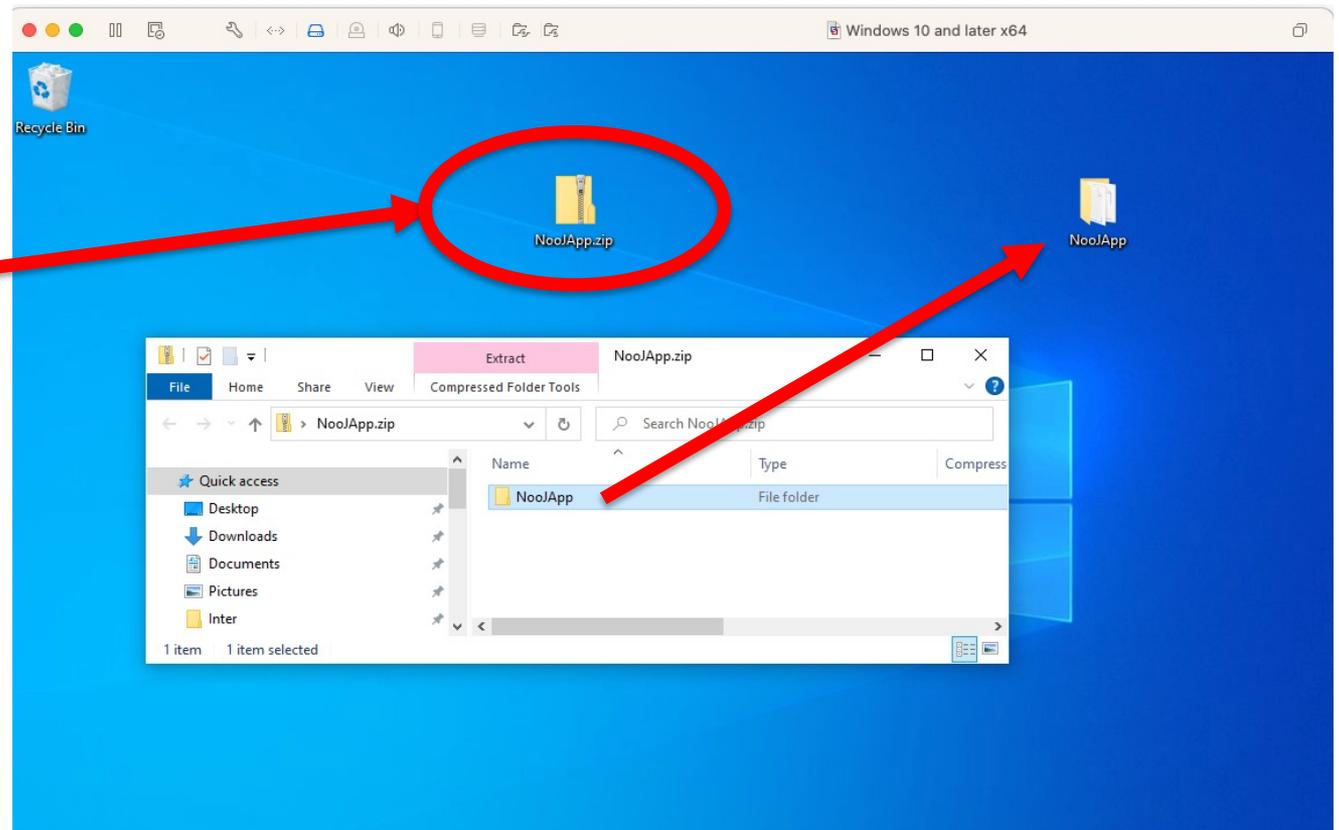
NooJ for Windows: Step 3

uncompress “NooJApp.zip”

- Double click here to open zip content

- Drag folder to Desktop

- NooJApp.zip is no longer necessary.



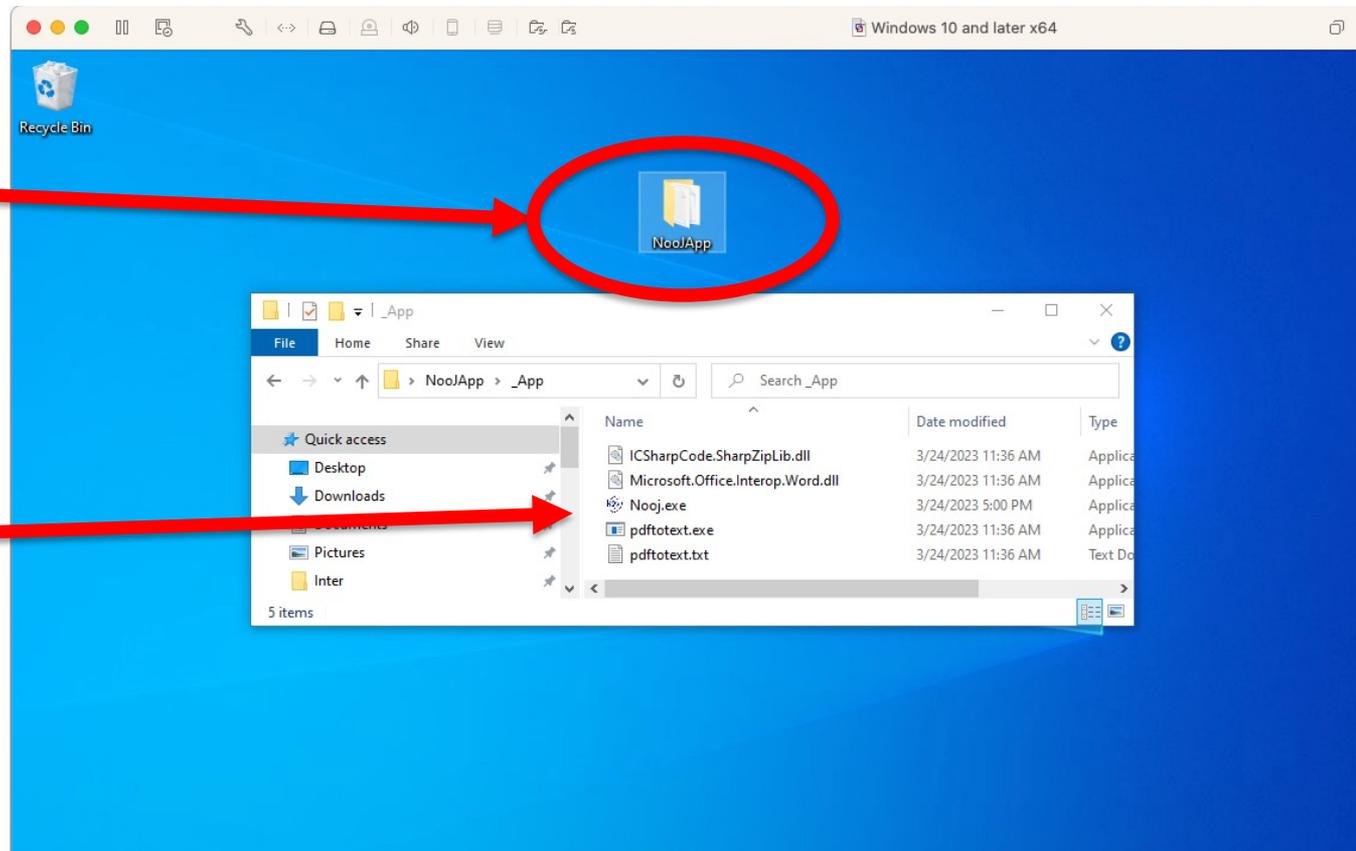
NooJ for Windows: Step 4

run Nooj.exe

- Double click here
to open the **NooJApp**
folder

- Open the **_App** folder

- Double click here
to run **Nooj.exe**



NooJ for Windows: Step 5

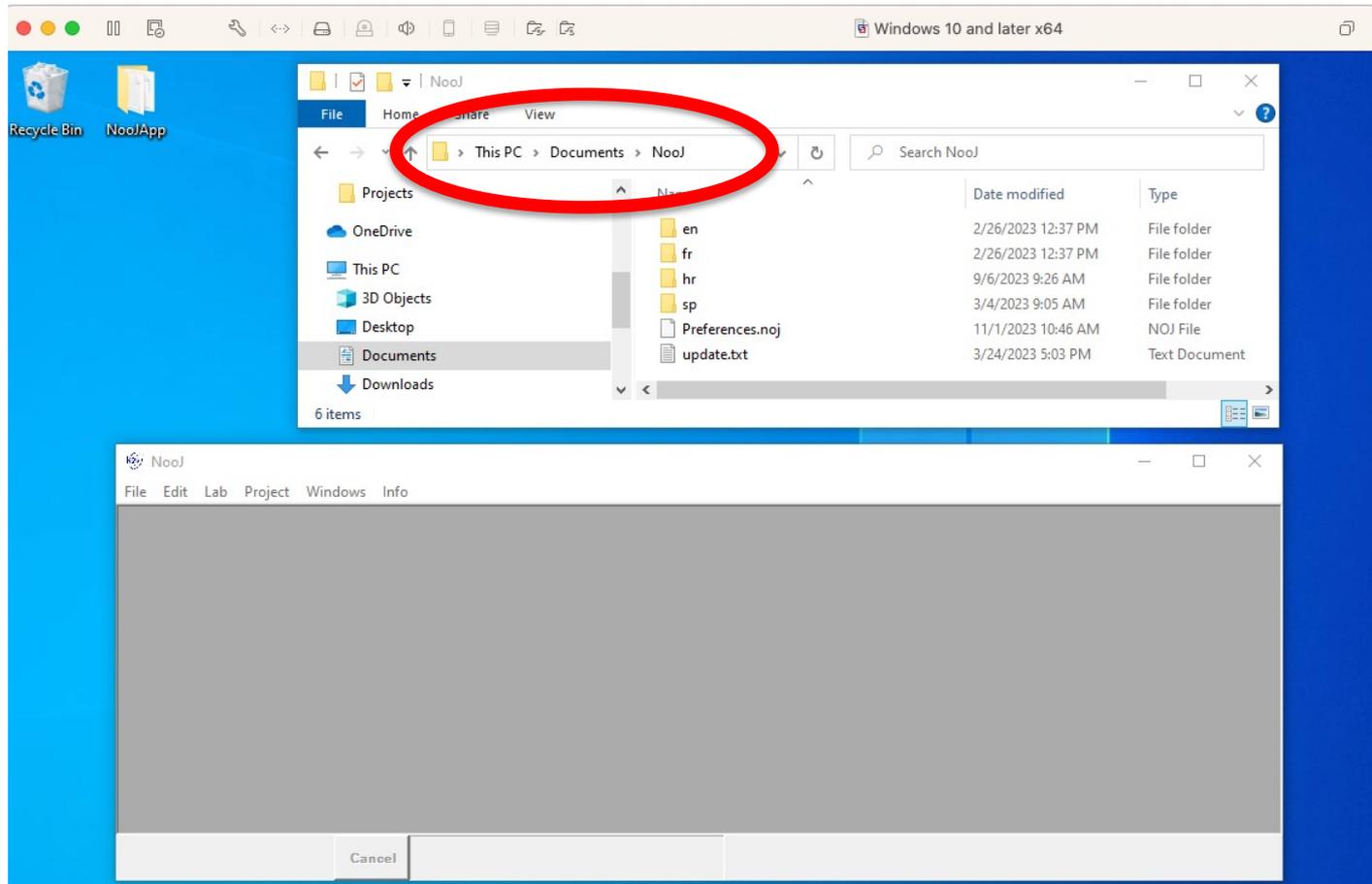
run Nooj.exe

- The first time NooJ runs, click “OK” to all questions. NooJ will create a folder **NooJ** in **Documents**

- the “en” folder contains resources for English

- the “fr” folder contains resources for French

- other resources can be downloaded from the NooJ Web site, page “Resources”



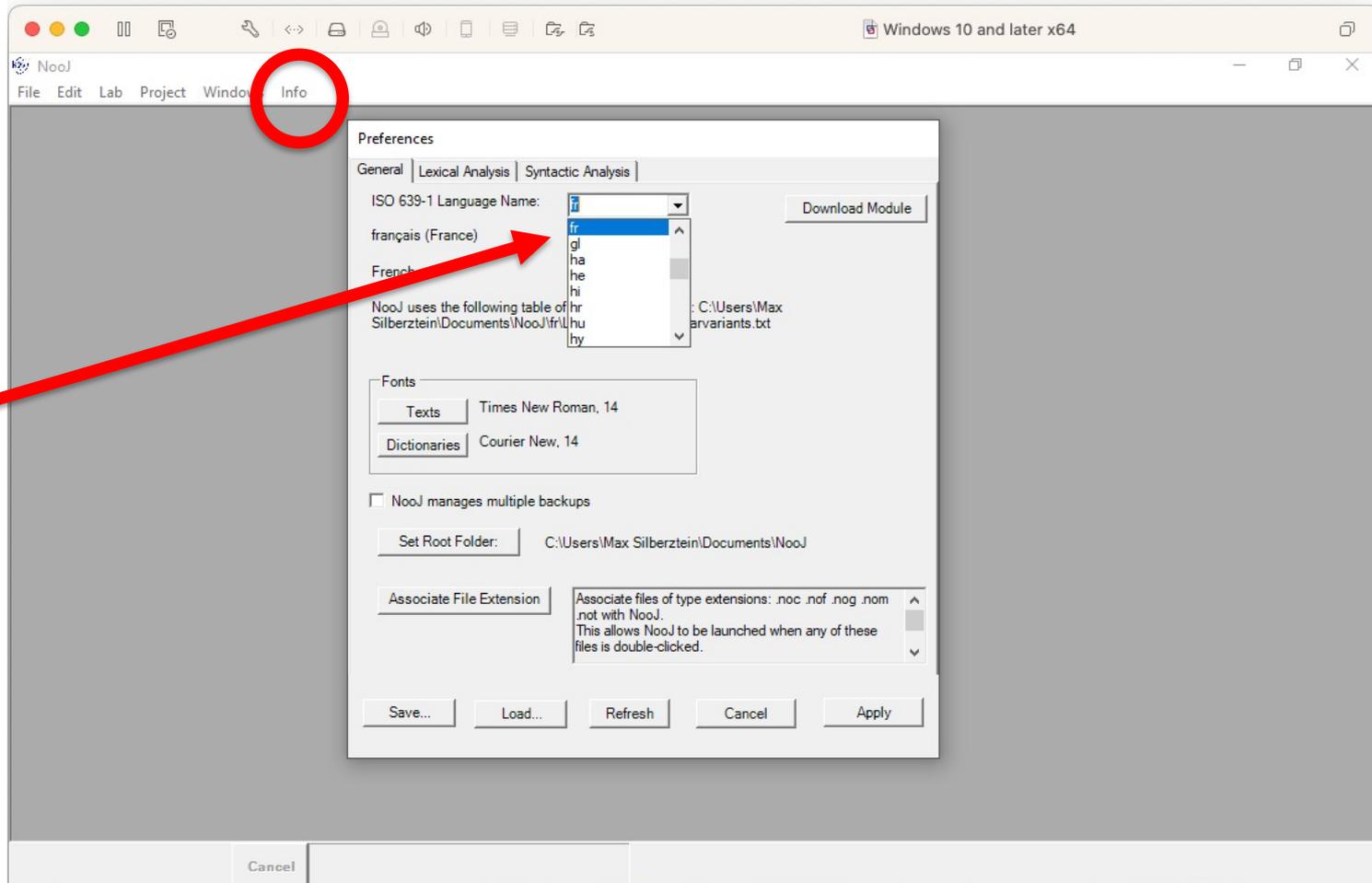
NooJ for Windows: Step 6

Set Default Language

- To set the default language, click **Info > Preferences**

- go to the **General** tab

- Select Language, e.g. "en" for English



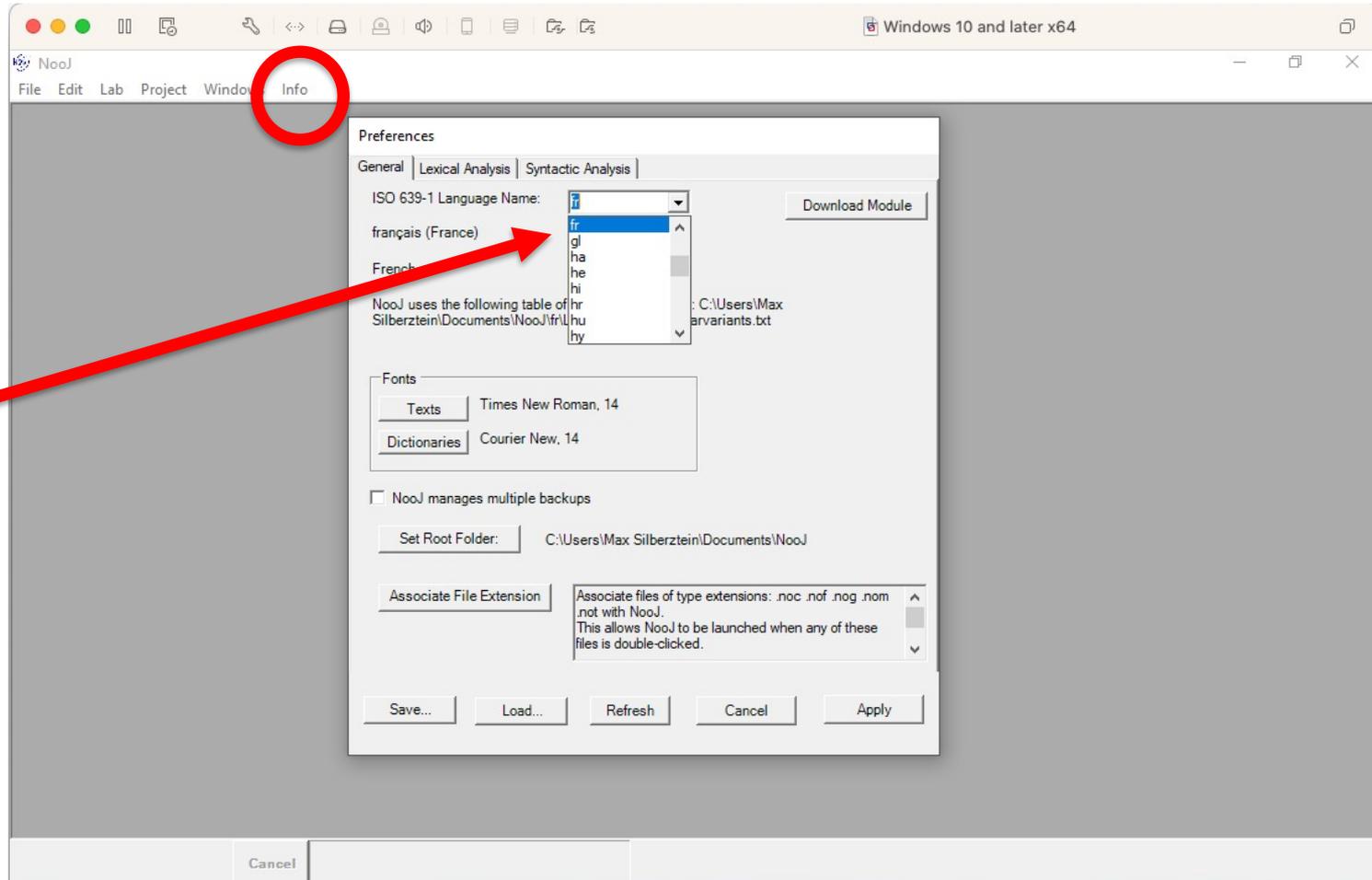
NooJ for Windows: Step 6

Set Default Language

- To set the default language, click **Info > Preferences**

- go to the **General** tab

- Select Language, e.g. "en" for English



NooJ for Windows: Step 7

Test NooJ

If the “en” language was selected:

- click **File > Open > Text**, then select “_The portrait of a lady” to open text.

- click **TEXT > Locate**

- Enter **women**

- click the red button to compute the concordance for wordform **women**

The screenshot shows the NooJ application window with the following components:

- Menu Bar:** File, Edit, Lab, Project, Window, Info, TEXT. The 'TEXT' menu is circled in red.
- Main Text Area:** Displays the text from the file "_The Portrait Of A Lady".
- Locate Dialog:** A dialog box titled "Locate a pattern in _The Portrait Of A Lady" with the following settings:
 - Pattern is: a NooJ regular expression:
 - Limitation: Only: 100 occ.
 - Reset Concordance
- Concordance Window:** A window titled "Concordance for text: _The Portrait Of A Lady.not" showing a table of results. A red button is visible at the bottom right of this window.

Text	Before	Seq.	After
	to his father. 'The pretty	women	themselves may be sent flying
	are rather inscrutable. They say	women	don't know how to
	had an excellent manner with	women	. The girl appreciated it in
	object of envy to many	women	; her light grey eyes, a
	world, but she had seen	women	who lied and who tried

Query 61/61

0.4 sec Cancel

Useful Information

- You might want to move the **NooJApp** folder from your desktop to another folder, *e.g.*, **c:\Program files**.
- You might need to manage read and execute rights so that other users of the PC can run NooJ.
- You might want to create a shortcut to the **Nooj.exe** file, so that you do not need to navigate to **NooJApp_App** everytime to run NooJ.
- Because NooJ processes several types of files (.dic, .noc, .nod, .nof, .nog, .nom, .not), you should set Windows File Explorer to display file name extensions.
- Do not mess with the content of the **NooJApp** folder: never move, edit, nor rename any file inside the **NooJApp** folder. Always work in your personal **Documents\NooJ** folder: all your files should be stored in your personal folder.
- Some NooJ file names start with the « _ » character. These files might get modified, updated or even deleted during a subsequent update of the NooJ software. If you need to modify them, save the modified files with another name that does not contain the « _ » prefix (preferably in you **Documents/NooJ** folder)

Problems?

- Newest versions of Microsoft Windows do not like software that do not come from Windows store.
 1. Open the Start menu, and click Settings.
 2. Click Apps on the left side, then click Advanced app settings.
 3. You should see a "Choose where to get apps" option. Collapse its drop-down menu, and select Anywhere.
- Windows says that .NET is not installed, or obsolete.
 - download latest version of .NET at: <https://dotnet.microsoft.com/en-us/download/dotnet-framework>
- An anti-virus does not like to download software from the internet.
 - temporarily unload antivirus.
- Remember: do not mess with the content of the NooJApp folder.
Always work in the **Documents\NooJ** folder
- Other problem?
 - Contact max.silberztein@gmail.com



CONGRATULATIONS



You have successfully installed

NooJ for Windows

(Now you have to read the documentation)

