

# NLP Suite

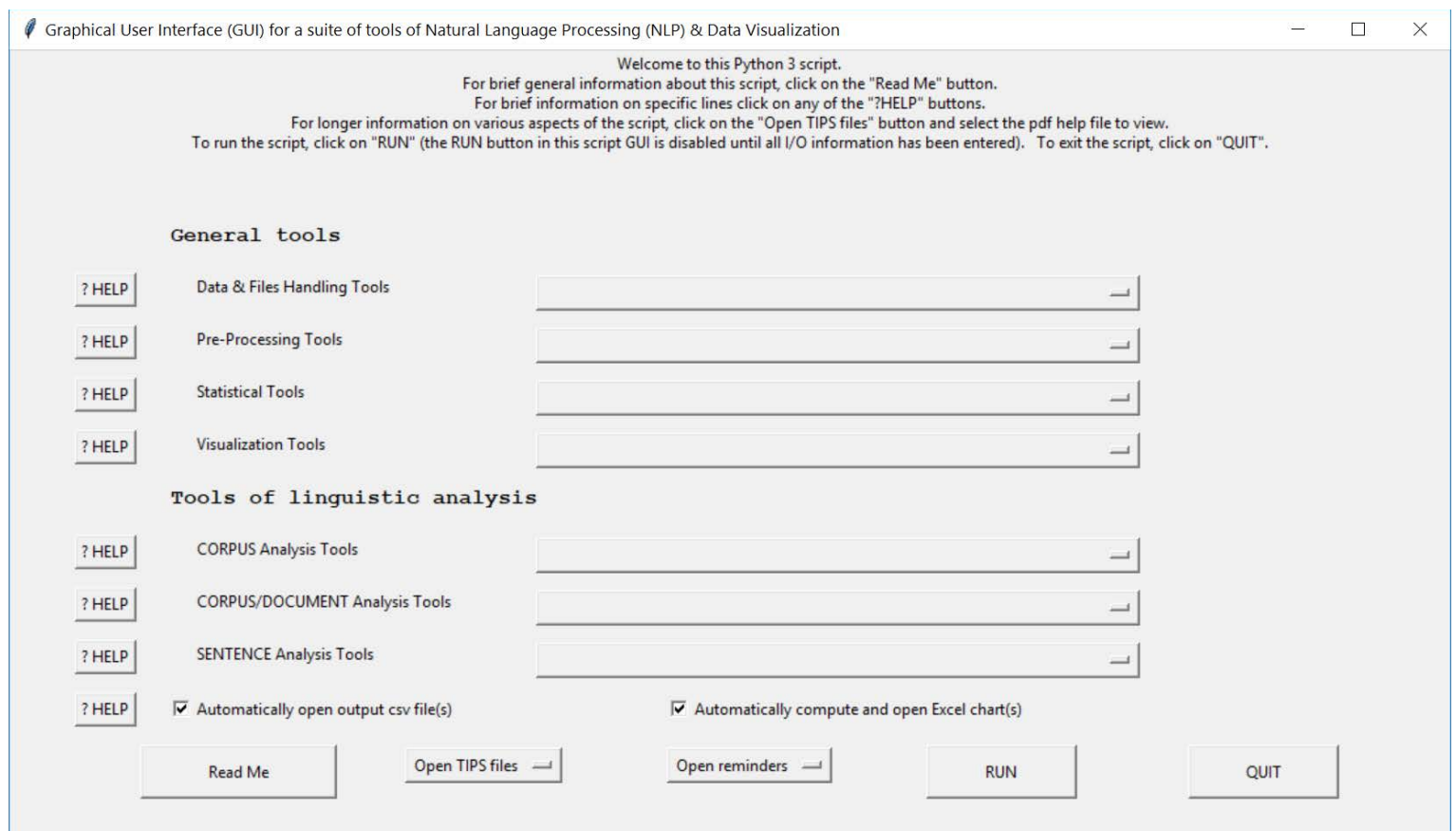
## Tools of Linguistic Analysis

### Table of contents

The NLP Suite main GUI (Graphical User Interface) .....	1
Tools of linguistics analysis.....	1
CORPUS Analysis Tools.....	2
CORPUS/DOCUMENT Analysis Tools .....	2
SENTENCE Analysis Tools .....	3

### The NLP Suite main GUI (Graphical User Interface)

When you fire up NLP\_main.py in command line, you will display the main NLP Suite GUI.



As the GUI makes clear, the NLP Suite provides two sets of tools: General tools and tools for linguistic analysis of texts via Natural Language Processing (NLP). This TIPS file provides a brief introduction to General tools. Read the TIPS\_NLP\_NLP Suite Tools of linguistic analysis.pdf.

**Select any of the tools using the dropdown menu and then click on RUN to open more specialized GUIs. The NLP GUI is only a front-end display of all the things you can do with the NLP Suite.**

### Tools of linguistics analysis

The label groups together a number of tools for dealing with the automatic computational analysis of texts, an approach better known as Natural Language Processing (NLP). The tools are divided into three broad classes depending upon what they require in input: a corpus of multiple documents, a corpus or a single file. And whatever

the input (corpus or single document) the Sentence Analysis Tools allow you to zoom into a more fine-grained analysis of texts at the sentence level.

### ***CORPUS Analysis Tools***

The tools in this group can ONLY be applied to a set of documents in input. They cannot be applied to a single document. Topic modeling with Mallet requires a further download and installation of the freeware Mallet.

Co-Occurrences viewer
N-grams viewer
Shape of stories
Similarities between documents (via Python difflib)
Similarities between documents (via Java Lucene)
Topic modeling (Gensim)
Topic modeling (Mallet)

### ***CORPUS/DOCUMENT Analysis Tools***

The tools under this label are the largest set of NLP tools in the Suite. At the heart of the group is the freeware, open source **Stanford CoreNLP** set of tools (parser, coreference, gender annotator, date and quote extractor). The Stanford CoreNLP parser produces in output the CoNLL table which is the basis of a large number of analytical tools in the NLP Suite (e.g., clausal analysis, noun and verb analysis). You will also find here such tools as Sentiment or Concreteness analysis or the SVO (Subject-Verb-Object extractor).

Stanford CoreNLP
Clausal analysis (via CoNLL)
Function words analysis (via CoNLL)
Noun analysis (via CoNLL)
Verb analysis (via CoNLL)
Narrative analysis
Style analysis
Concreteness analysis
Sentiment analysis
Date annotator (NER normalized date)
Gender annotator (male & female names)Gender guesser
N-grams (word & character)
Nominalization
Quote annotator (via CoreNLP)
Search CoNLL
Search text file(s) for n-grams & co-occurrences
Search text file(s) for words/collocations
Sentence complexity
Similarities between words (Levenshtein distance)
Spelling checker/Unusual words (via NLTK)
Spelling checker (via SpellChecker)
SVO extractor & visualization
Text annotator (via dictionary & dbpedia)
Text annotations extractor
Text readability (via textstat)
Word clouds

## ***SENTENCE Analysis Tools***

In this group, you will find tools that allow you to go deeper into the analysis of texts. Instead of broad, sweeping views of your corpus afforded by Topic modeling tools or n-grams Viewer, the sentence-level tools act more like a microscope rather than a telescope. By visualizing the behavior of various textual indicators by sentence index, the tools give you a way to explore the tempo of writing.

Sentence analysis (An overall GUI)  
Concreteness analysis by sentence index  
Dictionary items by sentence index  
Function words analysis by sentence index  
N-grams (word & character) by sentence index  
Noun analysis by sentence index  
Search words/collocations by sentence index  
Sentence complexity by sentence index  
Sentence visualization: Dependency tree viewer (png graphs)  
Sentence visualization: Dynamic sentence network viewer (Gephi graphs)  
Sentiment analysis by sentence index  
Text readability by sentence index (via textstat)