

Things to Do with Word NLP Approach

Table of Content

Let's start with a corpus: What is it?.....	2
Corpus of txt type files	2
Then... Let's move on to the questions: What do you want to know via <i>Distant Reading</i> ?	2
What is "Distant reading"?	2
My recommendation would be: Try a combination of close and distant reading!	2
A list of questions	2
A recommended sequence of analyses	3
Check the quality of input documents first!	3
Check utf-8 compliance	4
Change to utf-8 non-utf-8 apostrophes & quotes	4
Find & Replace String.....	4
Run spell checking	4
Get the lay of the land.....	5
What's in your corpus?	5
Visualizing word frequencies (word clouds)	8
n-grams, co-occurrences and the Viewer.....	9
Taking a closer look	9
Parse your corpus via Stanford CoreNLP.....	9
Become familiar with the CoNLL table: What is it?.....	9
Search the CoNLL table.....	10
Clausal analysis.....	10
Noun analysis	10
Verb analysis.....	10
Function words analysis.....	11
Dealing with time (CoreNLP normalized date)	11
Dealing with space	12
Annotate your corpus	12
What's in your corpus? Try DBpedia & YAGO for answers.....	13
Using a dictionary file	13
Are there people in your corpus and who are they, male or female? Gender annotator	14
Where exactly do annotations occur? At what point in each document?	14
WordNet	15
The web and flow, the tempo of writing: Sentence dynamic.....	15

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)
Verb analysis by sentence index
WordNet categories by sentence index

..... 16

Let's start with a corpus: What is it?

A corpus (from the latin *corpus*, body; plural *corpora*) is a collection of texts that you wish to analyze via NLP (e.g., the corpus of J.K. Rowling's Harry Potter novels).

Corpus of txt type files

In order to be processed by NLP tools, the **documents of your corpus must be in txt format** (not doc, docx, rtf, pdf...). If they are not txt, you can use one of the document type converters that the NLP Suite offers to convert the documents.

Then... Let's move on to the questions: What do you want to know via *Distant Reading*?

Let's start with a definition first, then move to a list of questions for a corpus.

What is "Distant reading?"

The idea of a "distant reading" goes back to Franco Moretti in his 2013 seminal work. It basically means, let the computer, rather than you (via "close reading") do the reading; let computer algorithms figure out what is going on in your corpus, particularly useful when dealing with very large corpora. Although...

My recommendation would be: Try a combination of close and distant reading!

A list of questions

And here is a list of questions you may wish to pose to your corpus.

1. How big is my corpus?
2. What is the average size of each document in the corpus?
3. Is the average document size evenly distributed or are there outliers that may affect analyses?
4. Which are the most frequent words in the corpus?
5. Which words tend to come together (collocations, such as "sit down" or "coming out")

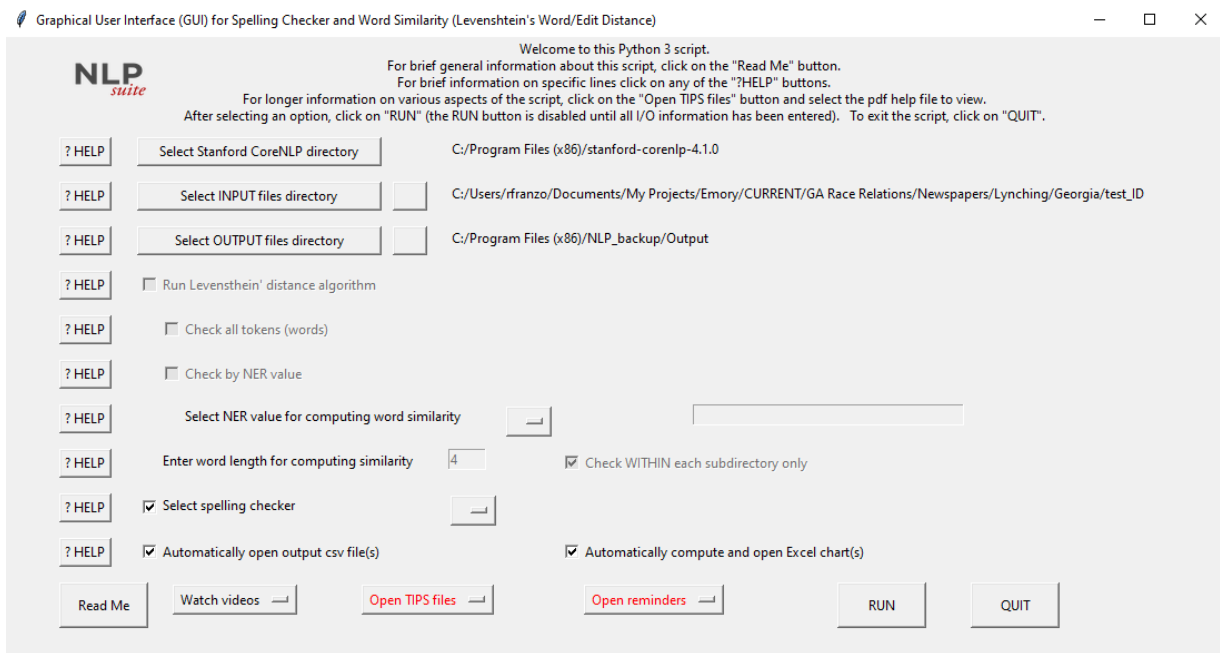
more often? And where in a text do they show up? At the beginning? In the middle? In the end? Everywhere?

6. Is there a way to grasp visually these statistics?
7. Are the documents about emotions, dialogues, movement?
8. Do the most frequent words change overtime at specific time points? For instance, do the most frequent words change with every new party leader in official Chinese government annual reports? Do they change with new chapters/books of Harry Potter?
9. Do certain events, as expressed in the words used by a document, fundamentally alter what we can expect the *before* and *after* in the text to be? For example, in people's health cancer narratives, what will be the fundamental differences between the words in the before and after the word "cancer" is used in the text? Similarly, in gay men stories, do documents split before and after "coming out" fundamentally talk about different things or in different ways?
10. What does the corpus talk about? What topics? Do different sets of documents talk about different topics?
11. Who is mentioned in a document? Men? Women? Organizations? Places? Times?
12. Do documents use direct speech, i.e., quotes? And who does the talking?
13. Are there specific locations mentioned in the documents, locations that can be placed on a geographic map?
14. If you think of a *scene* as place & time where characters act, how many scenes are there? How do they change?
15. And Who does What, When and Where?
16. And how can you analyze stories? Find the characters of a story? Its settings, as place and time?
17. And if the documents tell stories (e.g., novels or short stories), do these stories fundamentally fall into predictable "shapes"? (e.g., from rags to riches, from riches to rags, man in a hole)
18. Are the documents in the corpus easy to read? What level of education does it take to read them? How complex are the sentences? Does sentence complexity vary along a text?
19. How concrete/abstract is the language used?
20. What emotions do the texts try to evoke?
21. How are nouns, verbs, pronouns used? For instance, are verbs used in the present tense or past, or future? Or gerundive? In the active or passive voice? Are passive verbs more consistently used with specific (explicit or implicit) syntactic subjects?
22. Is there a tempo, a movement to the writing? Can such movement be visualized?
23. If the corpus consists of texts by different authors, are there differences in their use of language (by gender, race, nationality)?

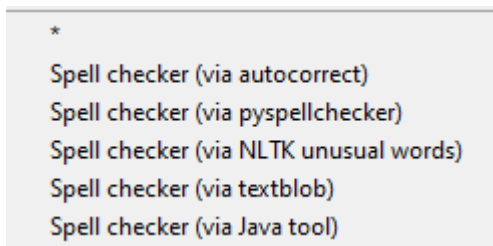
A recommended sequence of analyses

Check the quality of input documents first!

Run the script file_checker_converter_cleaner_main.py



Run any of the many available options for spell checking and automatically correct your input.



Get the lay of the land

Let's look at the tools that can provide a general overview of the corpus.

What's in your corpus?

The `whats_in_your_corpus_main.py` will display the GUI *A Sweeping View of Your Corpus*. This will give you several options to get a snapshot of the main characteristics of your corpus.

Graphical User Interface (GUI) for a Sweeping View of Your Corpus

Welcome to this Python 3 script.
For brief general information about this script, click on the "Read Me" button.
For longer information on specific lines click on any of the "?HELP" buttons.
After selecting an option, click on "RUN" (the RUN button is disabled until all I/O information has been entered). To exit the script, click on "QUIT".

? HELP Select Stanford CoreNLP directory C:/Program Files (x86)/stanford-corenlp-4.1.0

? HELP Select INPUT files directory C:/Users/rfranzo/Desktop/CORPUS DATA/Sample text

? HELP Select OUTPUT files directory C:/Program Files (x86)/NLP_backup/Output

? HELP ☒ Compute corpus statistics

? HELP ☒ What are the topics? (Topic modeling) ☐ via Mallet ☒ via Gensim

? HELP ☒ What else is in your corpus? (via Stanford CoreNLP and WordNet) *

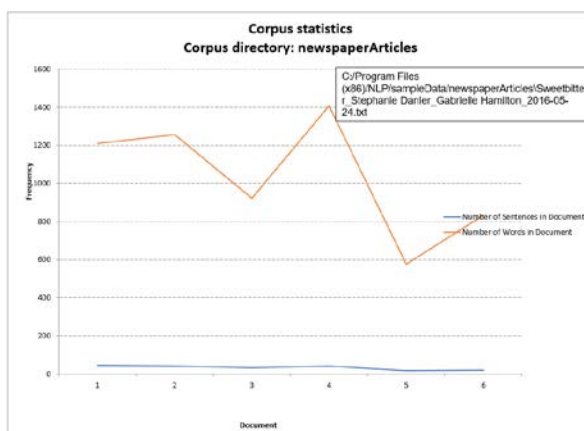
? HELP ☒ Automatically open output csv file(s) ☒ Automatically compute and open Excel chart(s)

Read Me Watch videos Open TIPS files Open reminders RUN QUIT

? HELP ☒ Compute corpus statistics

The *Compute corpus statistics* widget will give you basic corpus statistics: number of documents, sentences, and words per document, and frequency of top 20 words, and all visualized in Excel charts. Documents in the output csv file have hyperlinks that can open each document directly from the csv file.

Number of Documents	Document	Number of Sentences	Number of Words	Word1	Frequency	Word2	Frequency	Word3	Frequency	Word4	Frequency
6	1 C:/Program Files (x86)/NLP/sampleData/newspaperArticles/A Spool of Blue Thread_Anne Tyler_Rebecca Pepper Sinkler_2015-02-13.txt - Click once to follow. Click and hold to select this cell.	44	1208	1754 the	67	and	42	of	42	a	
6	2 C:/Program Files (x86)/NLP/sampleData/newspaperArticles/A Spool of Blue Thread_Anne Tyler_Rebecca Pepper Sinkler_2015-02-13.txt - Click once to follow. Click and hold to select this cell.	39	1208	74 the	63	and	32	to	29	a	
6	3 C:/Program Files (x86)/NLP/sampleData/newspaperArticles/A Spool of Blue Thread_Anne Tyler_Rebecca Pepper Sinkler_2015-02-13.txt - Click once to follow. Click and hold to select this cell.	74	1208	10 the	36	of	35	a	33	and	
6	4 C:/Program Files (x86)/NLP/sampleData/newspaperArticles/A Spool of Blue Thread_Anne Tyler_Rebecca Pepper Sinkler_2015-02-13.txt - Click once to follow. Click and hold to select this cell.	10	1208	37 the	98	and	45	of	43	a	
6	5 C:/Program Files (x86)/NLP/sampleData/newspaperArticles/A Spool of Blue Thread_Anne Tyler_Rebecca Pepper Sinkler_2015-02-13.txt - Click once to follow. Click and hold to select this cell.	37	1208	28 the	32	a	17	of	17	in	
6	6 C:/Program Files (x86)/NLP/sampleData/newspaperArticles/A Spool of Blue Thread_Anne Tyler_Rebecca Pepper Sinkler_2015-02-13.txt - Click once to follow. Click and hold to select this cell.	28	1208		45	to	25	a	23	and	

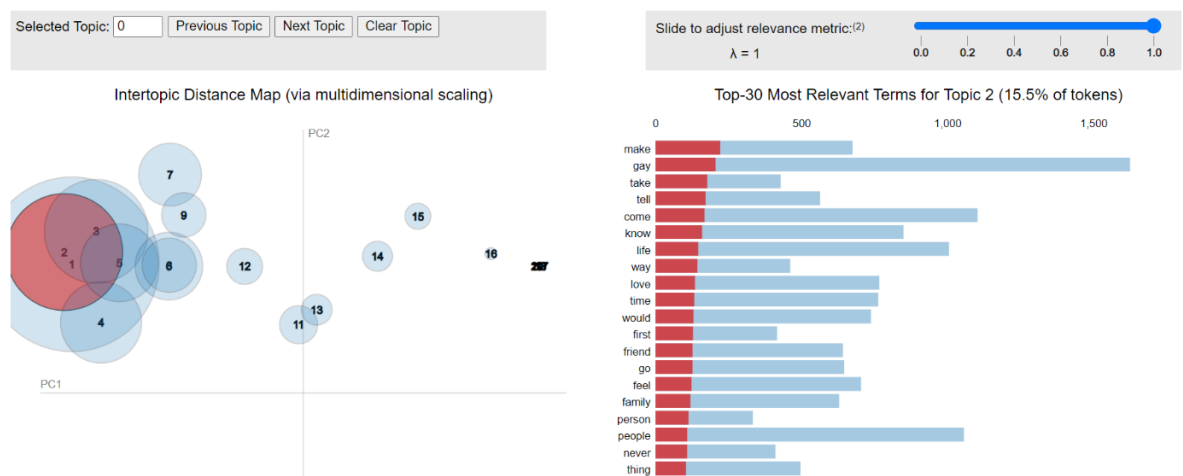


For each document, the algorithm also computes basic statistical measures for all computed frequencies and for each document (e.g., count, median, skewness, kurtosis).

Column header	Count	Mean	Mode	Median	Standard	Minimum	Maximum	Skewenes	Kurtosis	25% quan	50% quan	75% quantile	
Number of documents in corpus	36	6	0	6	0	6	6	0	0	6	6	6	
Document ID	21	3.5	0	1	3.5	1.870829	1	6	-1.2	-1.2	2.25	3.5	4.75
Number of Sentences in Document	198	33	0	42	37.5	11.89958	17	44	-1.99565	-1.99565	23.25	37.5	42
Number of Words in Document	6200	1033.333	0	575	1064.5	310.5599	575	1406	-1.0862	-1.0862	855	1064.5	1244.75
Number of Syllables in Document	8942	1490.333	0	837	1564	418.6658	837	1910	-0.83323	-0.83323	1264.5	1564	1817.75
Frequency1	341	56.83333	0	32	54	24.6042	32	98	0.382742	0.382742	38.25	54	66
Frequency2	196	32.66667	0	17	33.5	10.48173	17	45	-0.77246	-0.77246	26.75	33.5	40.25
Frequency3	187	31.16667	0	17	31	10.32311	17	43	-1.43346	-1.43346	24.5	31	39.75
Frequency4	167	27.83333	0	14	26.5	9.907909	14	41	-0.80526	-0.80526	22.75	26.5	34.75
Frequency5	129	21.5	0	14	19.5	6.442049	14	32	0.221977	0.221977	18.25	19.5	24.5

☒ What are the topics? (Topic modeling)
 ☐ via Mallet
 ☒ via Gensim

The *What are the topics* widget will run LDA topic modeling via Gensim. The output will give you a basic idea of the main topics dealt in your corpus. At a later point, you will want to run both Mallet and Gensim using their GUIs because you will have a much richer set of data in output.



Gensim topic modeling visualization of several hundred gay men stories

Try out the NLP Suite tools [topic_modeling_Gensim_main.py](#) and [topic_modeling_mallet_main.py](#).

TIPS_NLP_Topic modeling Gensim.pdf

TIPS_NLP_Topic modeling Mallet.pdf

TIPS_NLP_Topic modeling and corpus size.pdf

☒ What else is in your corpus? (via Stanford CoreNLP and WordNet)

The *What else is in your corpus* widget will answer by default a series of questions about your corpus using Stanford CoreNLP annotators and WordNet:

- ★
- Dialogues
- Noun and verb classes
- People (by gender) & organizations
- References to date & time
- References to geographical locations
- References to nature

1. Is there dialogue? And which documents in your corpus contain dialogues?
2. What are the noun and verb classes of your corpus (e.g, nouns of feeling, food, location, or verbs of communion, emotion, motion)
3. Are there people and organizations and are people males or females?
4. Are there standard references to date and time?
5. Are there references to geographical locations that be geocoded and mapped using the GIS_main GUI?
6. Are there references to nature? Animals, plants, weather, celestial bodies?

All of these questions

Visualizing word frequencies (word clouds)

One appealing way of visualizing word frequency is through word clouds where word colors and size can be used to express frequency. Some of the word clouds can be very artistic!



Word cloud of Subject-Verb-Object (SVOs, S in red, V in blue, O in green) from Murphy's *Six Miracles* sample story

Try out the many word cloud options in the NLP Suite tool **wordclouds_main.py**.

There are also other powerful ways of visualizing words in documents via HTML. Take a look at the later section on DBpedia, YAGO, dictionary annotators!

TIPS_NLP_Wordclouds Visualizing word clouds.pdf

n-grams, co-occurrences and the Viewer

You can compute the frequency of words (or even characters in words) that occur alone (1-gram or unigram), in pairs (2-gram or bigram), in sets of 3 (3-gram or trigram), 4... and so on. These will be visualized in Excel charts. You may even visualize them by sentence ID, i.e., plotting them in relation to where each type of n-gram occurs in the text. Some authors claim that n-grams, both characters and words, can be good indicators of an author's style.

The NLP Suite also comes with another way of visualizing n-grams, a tool very similar to the Google Ngram Viewer, <https://books.google.com/ngrams>. If your filenames embed a date (e.g., The New York Times_9-12-191_3_1), the NLP Suite n-gram viewer plots the *frequency of the user-selected sets of words over time* (via Excel line plots). Maybe some combination of words have gone up or down in their use overtime in the corpus. The n-gram viewer would tell you that.

Try out the NLP Suite tool **NGrams_CoOccurrences_Viewer_main.py**.

TIPS_NLP_Ngrams and Word co-occurrences VIEWER.pdf

TIPS_NLP_Google Ngram Viewer.pdf

TIPS_NLP_Ngrams (word & character).pdf

TIPS_NLP_Style analysis.pdf

Taking a closer look

Parse your corpus via Stanford CoreNLP

Natural language **parsers** are computer programs that work out the grammatical structure of sentences: which words in a text are nouns, verbs, adjectives? Which words are the subject or object of a verb? The Stanford CoreNP is one of the best freeware, open source parsers. Try out what it can do for you **Stanford_CoreNLP_main.py**.

TIPS_NLP_Stanford CoreNLP download install run.pdf

TIPS_NLP_Stanford CoreNLP parser.pdf

TIPS_NLP_Stanford CoreNLP OpenIE.pdf

Become familiar with the CoNLL table: What is it?

Parsers produce in output a file called CoNLL table that, for each word, registers its syntactical role and much more. The NLP Suite provides various tools for inspecting the information available in the CoNLL table.

TIPS_NLP_Stanford CoreNLP CoNLL table.pdf

TIPS_NLP_POSTAG (Part of Speech Tags) Stanford CoreNLP.pdf

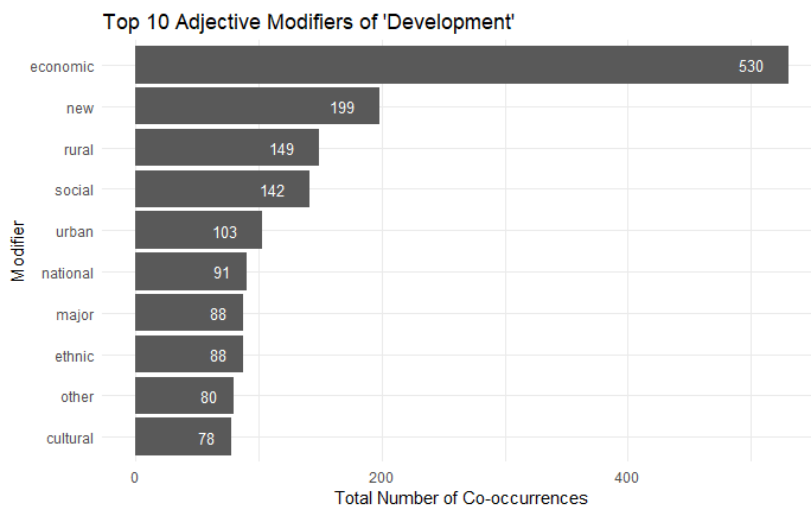
TIPS_NLP_DEPREL (Dependency Relations) Stanford CoreNLP.pdf

TIPS_NLP_NER (Named Entity Recognition).pdf

Search the CoNLL table

Try the **CoNLL_table_analyzer_main.py** to explore relationships between words by their syntactical roles.

For instance, “development” is one of the most frequent nouns in the annual Chinese government reports (2009-2019). But development about what? Linking adjectives to nouns gives us the following visual display.



Frequency of adjectives linked to the noun “development” in the annual Chinese government reports (2009-2019)

TIPS_NLP_NLP searches.pdf

Clausal analysis

What type of clauses are used in your corpus? Find an answer by using the NLP Suite tool **CoNLL_table_analyzer_main.py**.

TIPS_NLP_Clausal analysis.pdf

Noun analysis

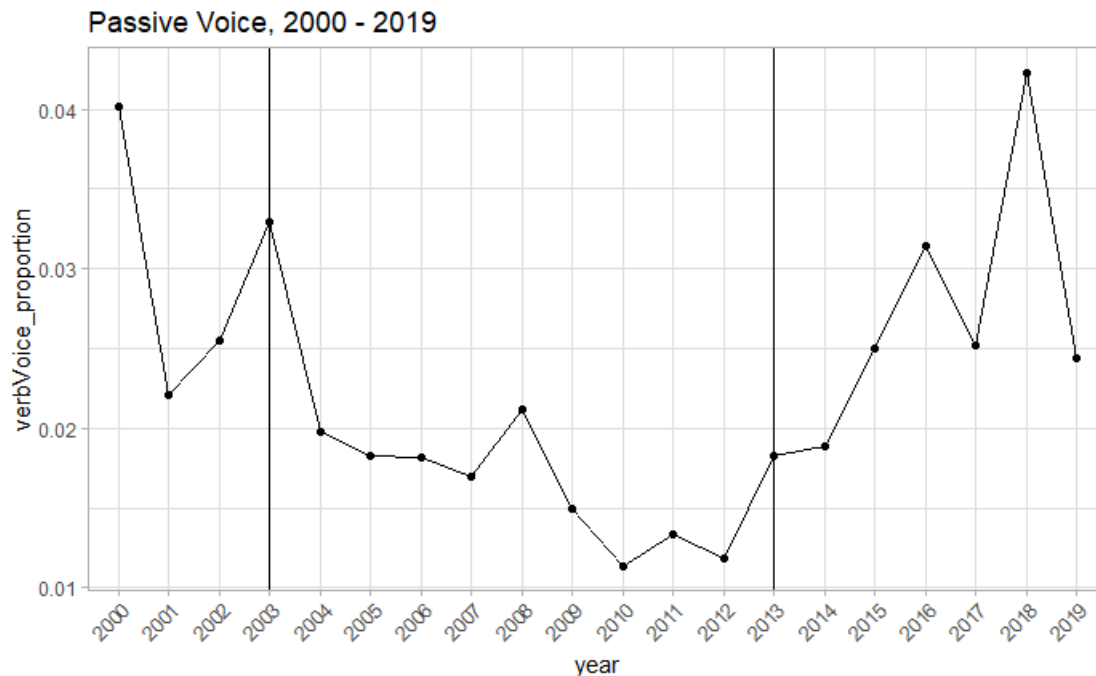
What types of nouns are used in your corpus? Proper nouns, like Chicago, or improper nouns, like city? How many such different nouns are used? Find the answers with **CoNLL_table_analyzer_main.py**.

TIPS_NLP_Noun Analysis.pdf

Verb analysis

Verbs are even more complicated than nouns. They have *voice* (active or passive), they express *modality* (e.g., can, will, must), *tense* (e.g., present or past). Find out how verbs behave in your corpus running **CoNLL_table_analyzer_main.py**.

The interesting questions about verbs is: do they change with specific authors, specific historical events? Interestingly, verb voice changes with every new Chinese leadership as reflected in the annual government reports (2009-2019).



TIPS_NLP_Verb Analysis.pdf

Function words analysis

Function words are pretty amazing little guys, such as articles (e.g., the, a), pronouns (e.g., she, they), auxiliaries (e.g., have, be). They constitute the vast majority of words in a text. Typically, they are called “stop words,” because normally they are “stopped” from being processed. They are simply thrown out from the analyses of texts. Yet, some scholars believe that these function words provide important clues about authors’ style, their gender, their mental state. Definitely worth exploring what these guys have to offer via

CoNLL_table_analyzer_main.py.

TIPS_NLP_Function Words Analysis.pdf

Dealing with time (CoreNLP normalized date)

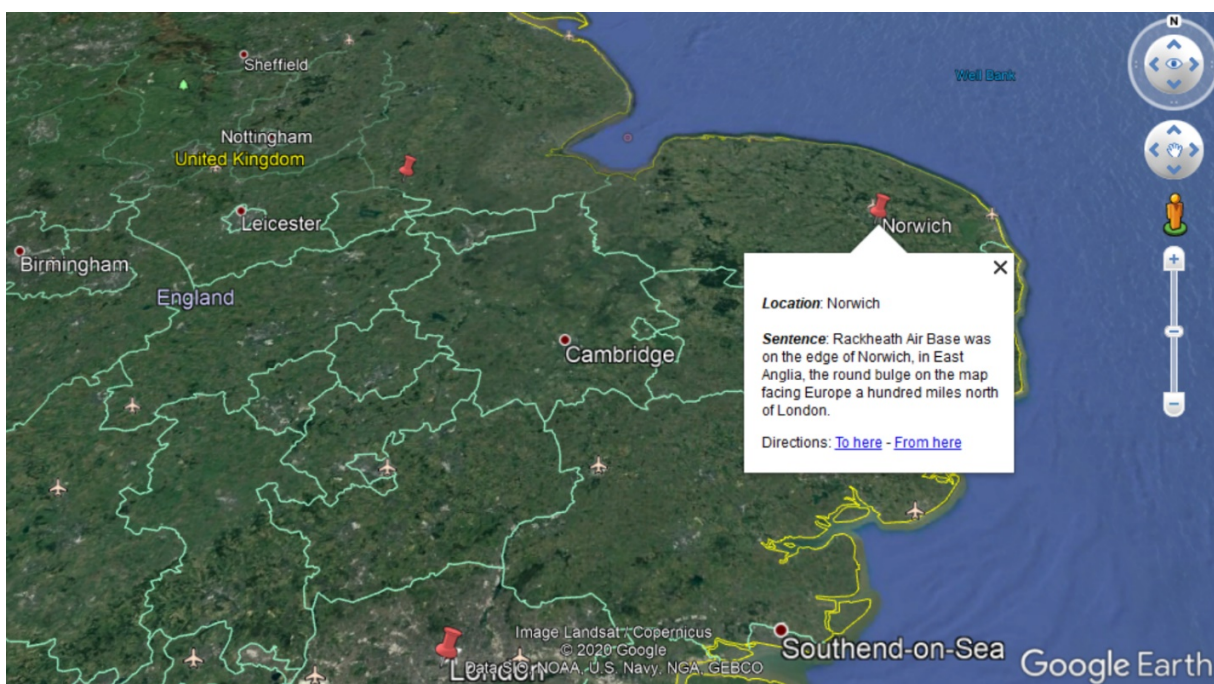
If your corpus is about scientific papers, time may not matter. But if your corpus is about stories, then time and space are fundamental categories of narrative. The Stanford CoreNLP normalized time annotator will standardize any temporal reference found in the text in a way that these can be plotted for their frequency, sorted from past to future. The use of time across a story will allow you to address the narrative typology of story and plot, of the chronological order of events (story) and their temporal re-ordering for rhetorical purposes to grab the reader (plot), were, like in a Hitchcock’s movie, you start from the end (the murder) and then reconstruct the story.

You will find the NER normalized date extractor in **Stanford_CoreNLP_main.py**. Check it out!

TIPS_NLP_Stanford CoreNLP date extractor.pdf

Dealing with space

The NLP Suite comes with a set of tools that allow you 1. to extract any reference to a special location found in the text (e.g., London) and 2. geocode this location, i.e., find its latitude and longitude; and 3. put these locations on a map, perhaps even a dynamic map if your locations are connected to specific dates/times so that the display of locations on the map moves interactively with time.



Google Earth map of geographic locations in Murphy's *Six Miracles* sample story
The NLP Suite tool that makes possible the seamless flow from texts to maps is **GIS_main.py**. By all means, try it out!

TIPS_NLP_Geocoding.pdf

TIPS_NLP_Google Earth Pro From KML to Excel.pdf

TIPS_NLP_Google Earth Pro.pdf

Annotate your corpus

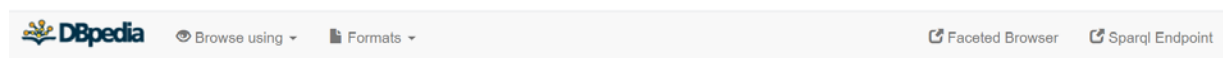
Annotators of various kind are available in the NLP Suite to give you further visual tools to aid you figure out what's going on in your corpus. An annotator will tag specific words or expressions in different colors or weight e.g., bold). You can annotate using DBpedia, basically Wikipedia or using a dictionary list of some kind (e.g., aggregate values taken from Wordnet or location names or by sentiment analysis values). You can also annotate the first names of people in the corpus by the gender, male or female, of the name.

What's in your corpus? Try DBpedia & YAGO for answers

You can use such knowledge-based databases as DBpedia or YAGO to annotate anything of general value expressed in the corpus. Most likely, DBpedia or YAGO will use Wikipedia entries to figure out if there is anything worthwhile annotating in your corpus. William Faulkner's short story *Dry September* annotated using the default classes will look like this:

[WILLIAM FAULKNER](#) Dry September THROUGH THE BLOODY September twilight, aftermath of sixty-two rainless days, it had gone like a fire in dry [grass](#): the rumor, the story, whatever it was. Something about Miss Minnie Cooper and a Negro. Attacked, insulted, frightened: none of them, gathered in the barber shop on that Saturday evening where the [ceiling fan](#) stirred, without freshening it, the vitiated air, sending back upon them, in recurrent surges of stale [pomade](#) and lotion, their own stale breath and odors, knew exactly what had happened. Except it wasn't Will Mayes, a barber said. He was a man of middle age; a thin, sand-colored man with a mild face, who was [shaving](#) a client. I know Will Mayes. He's a good [nigger](#). And I know Miss Minnie Cooper, too. What do you know about her? a second barber said. Who is she? the client said. A young girl? No, the barber said. She's about forty, I reckon. She ain't married. That's why I don't believe. Believe, hell! a hulking youth in a sweat-stained [silk](#) shirt said. Wont you take a white woman's word before a [nigger's](#)? I don't believe Will Mayes did it, the barber said. I know Will Mayes. Maybe you know who did it, then. Maybe you already got him out of town, you damn nigger-lover. I don't believe anybody did anything. I don't believe anything happened. I leave it to you fellows if them ladies that get old without getting married don't have notions that a man can't. Then you are a hell of a white man, the client said. He moved under the cloth. The youth had sprung to his feet. You don't? he said. Do you accuse a white woman of lying? The barber held the [razor](#) poised above the half-risen client. He did not look around. It's this durn weather, another said. It's enough to make a man do anything. Even to her. Nobody laughed. The barber said in his mild, stubborn tone: I ain't accusing nobody of nothing. I just know and you fellows know how a woman that never You damn nigger-lover! the youth said. Shut up, [Butch](#), another said. We'll get the facts in plenty of time to act. Who is? Who's getting them? the youth said. Facts, hell! You're a fine white man, the client said. Ain't you? In his frothy beard he looked like a [desert rat](#) in the moving pictures. You tell them, [Jack](#), he said to the youth. If there ain't any white men in this town, you can count on me, even if I ain't only a drummer and a stranger. That's right, boys, the barber said. Find out the truth first. I know Will Mayes. Well, by [God](#)! the youth shouted. To think that a white man in this town. Shut up, [Butch](#), the second speaker said. We got plenty of time. The client sat up. He looked at the speaker. Do you claim that anything excuses a [nigger](#) attacking a white woman? Do you mean to tell me you are a white man and you'll stand for it? You better go back North where you came from. The South don't want your kind here. North what? the second said. I was born and raised in this town. Well, by [God](#)! the youth said. He looked about with a strained, baffled gaze, as if he was trying to remember what it was he wanted to say or to do. He drew his sleeve across his [sweating](#) face. Damn if I'm going to let a white woman. You tell them, [Jack](#), the drummer said. By [God](#), if they. The screen door crashed open. A man stood in the floor, his feet apart and his heavy-set body poised easily. His white shirt was open at the throat; he wore a felt hat. His hot, bold glance swept the group. His name was McLendon. He had commanded troops at the front in [France](#) and had been decorated for valor. Well, he said, are you going to sit there and let a black son [rape](#) a white woman on the streets of [Jefferson](#)? [Butch](#) sprang up again. The [silk](#) of

Clicking on any of the annotated entries, for example on WILLIAM FAULKNER, will take us to the DBpedia website



William Cuthbert Faulkner (/ˈfɔːlkənər/, September 25, 1897 – July 6, 1962) was an American writer and Nobel Prize laureate from Oxford, Mississippi. Faulkner wrote novels, short stories, a play, poetry, essays, and screenplays. He is primarily known for his novels and short stories set in the fictional Yoknapatawpha County, based on Lafayette County, Mississippi, where he spent most of his life.

TIPS_NLP_Annotator DBpedia.pdf
TIPS_NLP_Annotator DBpedia ontology classes.pdf
TIPS_NLP_Annotator YAGO (schema.org) ontology classes.pdf
TIPS_NLP_Annotator.pdf
TIPS_NLP_Annotator extractor.pdf
TIPS_NLP_Gender annotator.pdf

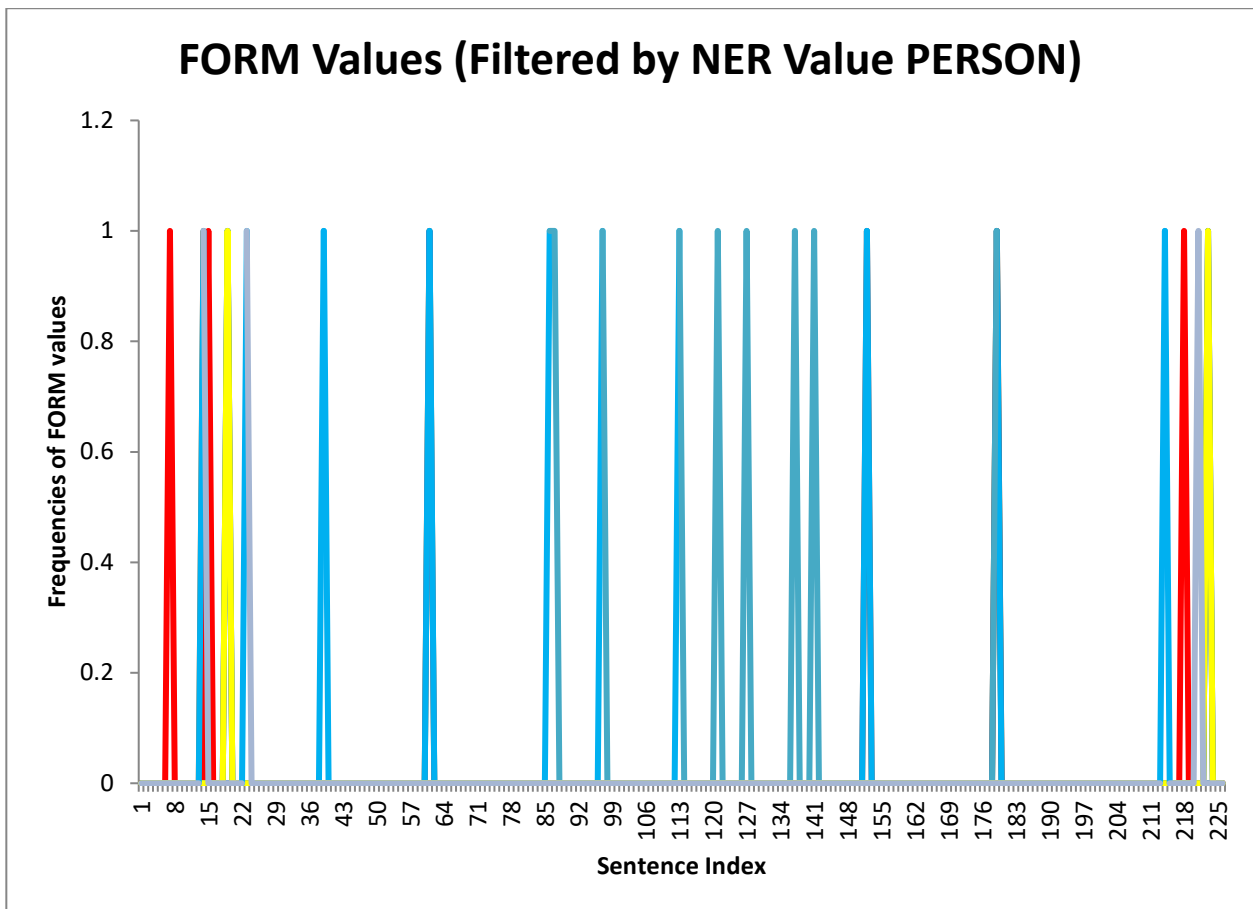
Using a dictionary file

Suppose that you have a listing of people in a University department of computer science. That is your dictionary. Now, you have a corpus of scholarly articles and you want to see who gets cited in those articles using your list. Or, you have a list of cities, of students, of prime ministers in a country. Or you simply have your corpus, you extract all the locations mentioned in the corpus (NER location values) and then you use these values to tag your corpus.

TIPS_NLP_Annotator.pdf
TIPS_NLP_Annotator dictionary.pdf

Are there people in your corpus and who are they, male or female? Gender annotator

The NLP Suite provides two different approaches to gender annotation, that is, to annotating text based on the mentions of people by their gender: 1. The Stanford CoreNLP gender annotator; 2. A dictionary-based annotator using several different kind of lists (e.g., US census names, US Social Security, or Carnegie Mellon lists). The tool further allows you to visualize where names are used: do all male names cluster together on one part of the document and female names in another or do they mix together? Do certain names come up at different points in the text?



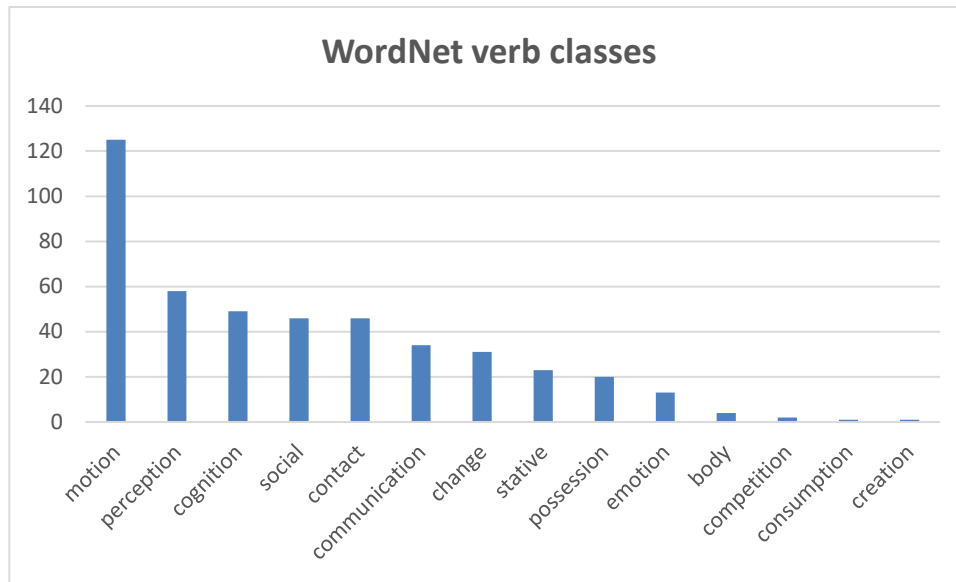
Line plots of frequency of personal nouns by sentence index (blue/red/yellow for male/female/neutral first names) in Murphy's *Six miracles* sample story

Where exactly do annotations occur? At what point in each document?

You can annotate (tag) a text and then display it to see what is going on. But you can then extract the tags to provide both raw statistics and visual charts of their use.

WordNet

WordNet is a huge lexical database developed at Princeton University of over 100,000 words (nouns, verbs, adjectives, and adverbs) and widely used in NLP projects. You can use WordNet verb categories, for instance, to find out how the verbs in a document broadly classify. For instance, most verbs in the Murphy *Six Miracles* sample text distributed with the NLP Suite come under the heading of verbs of motion.



Bar chart of frequency distribution of WordNet verb classes in Murphy's *Six Miracles* sample story

What is of interest is not just the distribution of word classes but *where* certain classes of verbs appear in a text. Do those verbs of motion cluster at specific points in the text or are they more or less evenly distributed throughout the text? Find the answers to these questions using the NLP Suite tool **WordNet_main.py**.

TIPS_NLP_WordNet.pdf

The web and flow, the tempo of writing: Sentence dynamic

Writing, particularly good writing, is not as flat as a billiard table. It has a tempo that comes and goes, sentence after sentence, reflected in each sentence properties, from sentence length, sentence complexity, its sentiment, concreteness, characters, time and locations. Every aspect of a sentence contributes to the overall tempo of writing. Zoom in at that level and visualize each sentence in the dynamic context of other sentences. In the NLP GUI under SENTENCE Analysis Tools you will find the many available options to visualize the tempo of writing. Try out **NLP_main.py** or **sentence_analysis_main.py**.

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)
Verb analysis by sentence index
WordNet categories by sentence index