

The ABC of Computational Text Analysis

#10 NLP WITH PYTHON

Alex Flückiger
Faculty of Humanities and Social Sciences
University of Lucerne

11 May 2024

Recap last Lecture

- perform first real-world (data) analysis 
analyse discourse on wokeness in Swiss media

Outline

- get some organizational stuff done
- let's do serious NLP! 🌟
- code interactively

interrupt, ask, and complement

Organizational



Course Evaluation

Tell me...



Please follow the link in the email, received on 8 May 2024

Thanks for any constructive feedback,
be it sweet or sour!



~~Assignment #3~~

No further assignments!

- Focus on the mini-project!
- Record your projects [here](#) by 18 May!

Issues on Mac

ModuleNotFoundError: No module named `plotnine`

- the `issue` is the consequence of a failed `gensim` installation
- navigate to project folder `KED2024` in Terminal
- run the following commands

```
poetry remove scattertext
```

```
poetry install
```

A Primer on classic NLP

What is a Word?

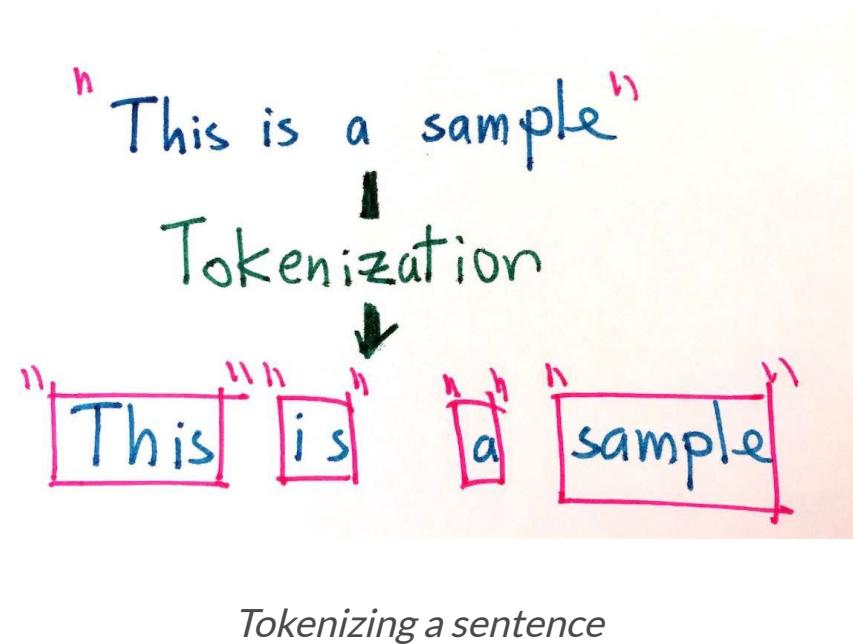
- words ~ segments between whitespace
- yet, there are ...

contractions: U.S., don't

collocations: New York

Token

- token ~ computational unit
representation of words
- lemma ~ base form of a word
texts → *text*
goes → *go*
- stop words ~ functional words
lacking deeper meaning
the, a, on, and ...



Let's tokenize this sentence! Isn't it easy? 😊

Common Processing Steps in NLP

1. Tokenizing

segmenting text into words, punctuations etc.

2. Tagging part-of-speech (POS)

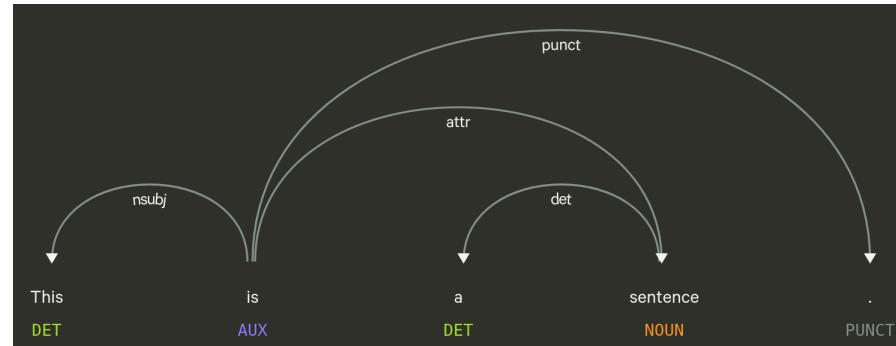
assigning word types (e.g. verb, noun)

3. Parsing

describing syntactic relations

4. Named Entity Recognition (NER)

organizations, persons, locations, time etc.



Catch up on NLP with

Jurafsky and Martin (forthcoming)



Questions?

References

Jurafsky, Dan, and James H. Martin. forthcoming. *Speech and Language Processing*. 3rd (Jan 7, 2023 draft). London: Prentice Hall. <https://web.stanford.edu/~jurafsky/slp3/>.