Lemmatization

Ir order to analyze a word, it is useful to identify its lemma, which is the inflected form or dictionary form. Lemmatization is the algorithmic process of identifying the lemmata of words.

We start by importing spaCy and the language library. Then as usual, we define our text. The lemmatization function is token.lemma_ and we apply it as follows:

```
for token in doc:
    print(token.text, token.lemma_)
```

Go ahead and check the notebook of this tutorial to see how to define a function.

Stop words

In computational linguistics, sometimes we want to filter out the most common words in a language, if these words do not have relevance for the analysis. Stop words are for example articles ('an', 'a', 'the'), conjunctions (e.g. 'for', 'and', 'but') and prepositions (e.g. 'in', 'under', 'before').