Moduldokumentation

Modul Natural Language Processing and Probability (nlp)

Simon Wächter

2018

Inhalt

[1 Einleitung 2](#_Toc507831755)

[1.1 Einleitung 2](#_Toc507831756)

[1.2 Lernziele 2](#_Toc507831757)

[1.3 Prüfungen 2](#_Toc507831758)

[2 Woche 1 3](#_Toc507831759)

[2.1 Lerninhalte 3](#_Toc507831760)

[2.2 Abgrenzung zu ML und Deep NLP 3](#_Toc507831761)

[2.3 Lernziele Woche 1 3](#_Toc507831762)

[2.4 What is Natural Language Processing 4](#_Toc507831763)

[2.5 Short Exercise 4](#_Toc507831764)

[2.6 Modern NLP Applications 4](#_Toc507831765)

[2.7 Communication Problems 5](#_Toc507831766)

[2.8 Typical NLP Pipeline 5](#_Toc507831767)

[2.9 Example: Email & Chatbot Pipeline 6](#_Toc507831768)

[2.10 Genres of Text 6](#_Toc507831769)

[2.11 Two Examples 7](#_Toc507831770)

[2.12 Praktischer Teil 7](#_Toc507831771)

[3 Woche 2 8](#_Toc507831772)

[3.1 Lernziele Woche 2 8](#_Toc507831773)

[3.2 Part-of-Speech 8](#_Toc507831774)

[3.3 Typical NLP Pipeline 8](#_Toc507831775)

[3.4 Part-of-Speech 9](#_Toc507831776)

[3.5 Tagset 9](#_Toc507831777)

[3.6 Tagged Corpora 11](#_Toc507831778)

[3.7 Part-of-Speech 11](#_Toc507831779)

[3.8 Part-of-Speech distribution 12](#_Toc507831780)

[3.9 The POS problem 13](#_Toc507831781)

[4 Woche 3 14](#_Toc507831782)

# Einleitung

## Einleitung

Dieses Dokument stellt die Moduldokumentation für das Modul nlp dar. Allfällige Unterlagen sind im Modulordner zu finden.

## Lernziele

Das Modul beinhaltet folgende Lernziele:

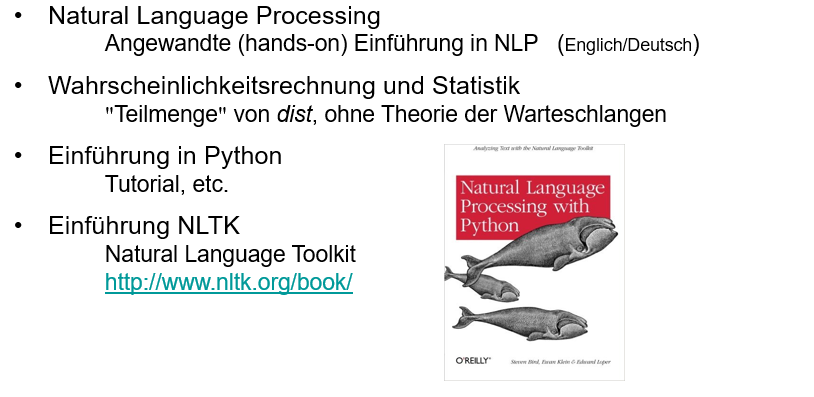
* Die Studierenden kennen die Grundlagen des Natural Language Processing und die wichtigsten Anwendungsgebiete.
* Die Studierenden können wichtige Methoden der Wahrscheinlichkeitsrechnung und Statistik anwenden.
* Sie sind in der Lage, NLP-Algorithmen zu implementieren und auf Texte in Deutscher und Englischer Sprache anzuwenden.
* Die Studierenden können die gewählten Modelle und Resultate quantitativ analysieren, bewerten und interpretieren.
* Die Studierenden können mit grossen Text-Corpora umgehen und eine geeignete NLP-Toolbox einsetzen.

## Prüfungen

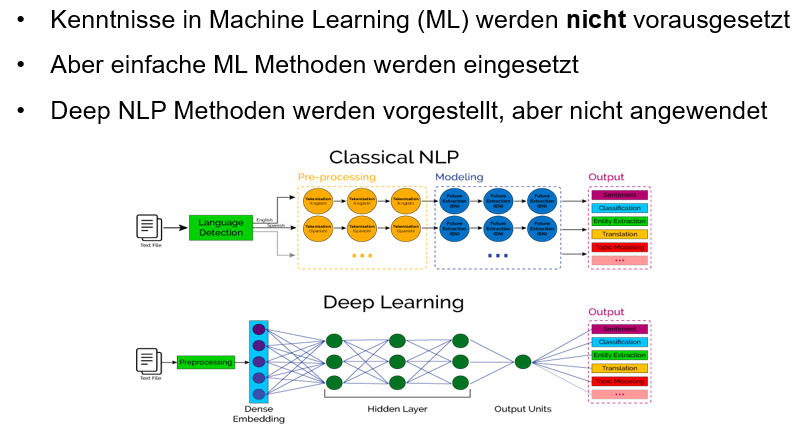
Die Modulnote setzt sich zu 5 Assessments zu 50%, einem Vortrag zu 10% und 2 Prüfungen zu 40% zusammen. Zudem müssen mindestens 9 Vorträge besucht werden.

# Woche 1

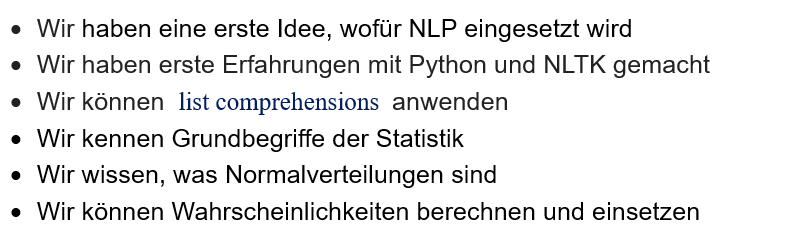
## Lerninhalte



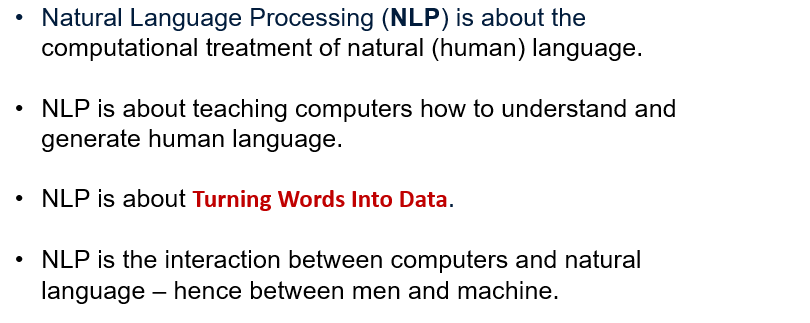
## Abgrenzung zu ML und Deep NLP



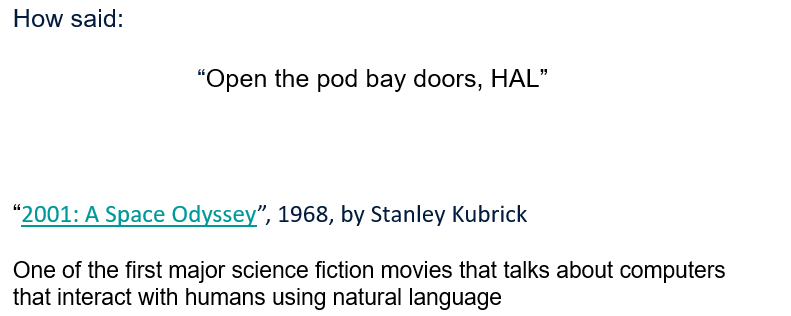
## Lernziele Woche 1



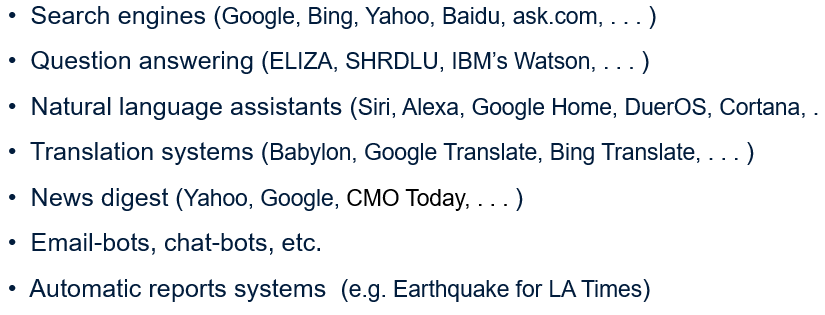
## What is Natural Language Processing



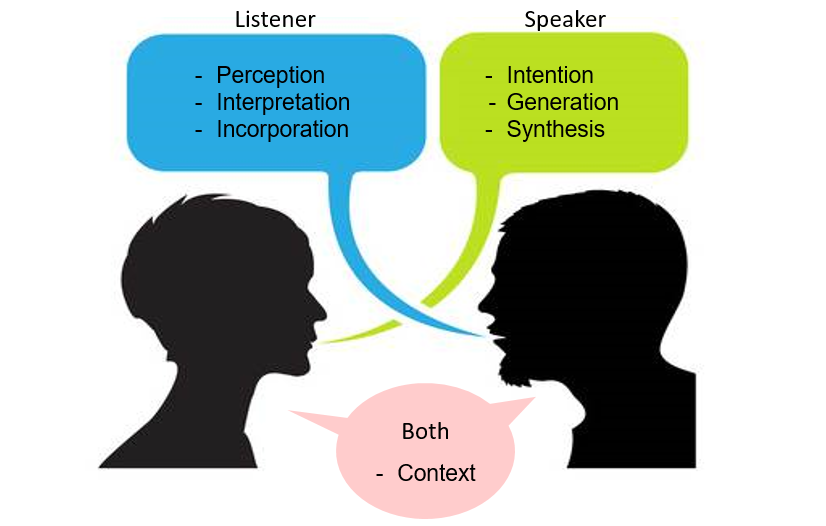
## Short Exercise



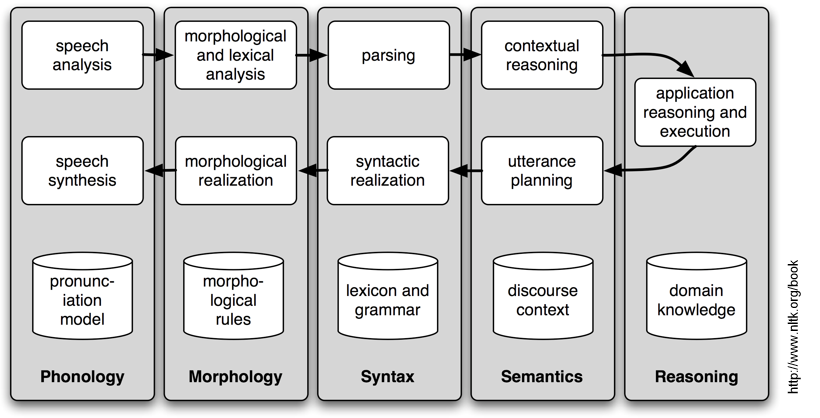
## Modern NLP Applications



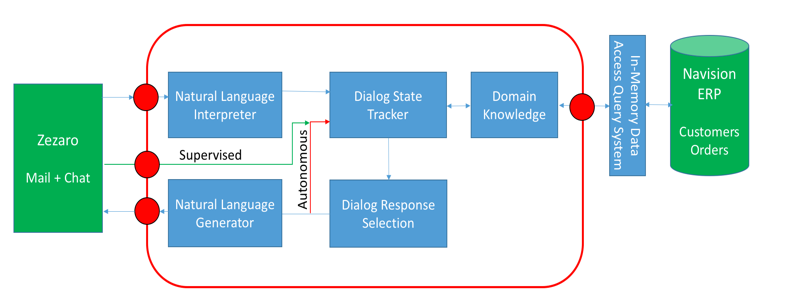
## Communication Problems

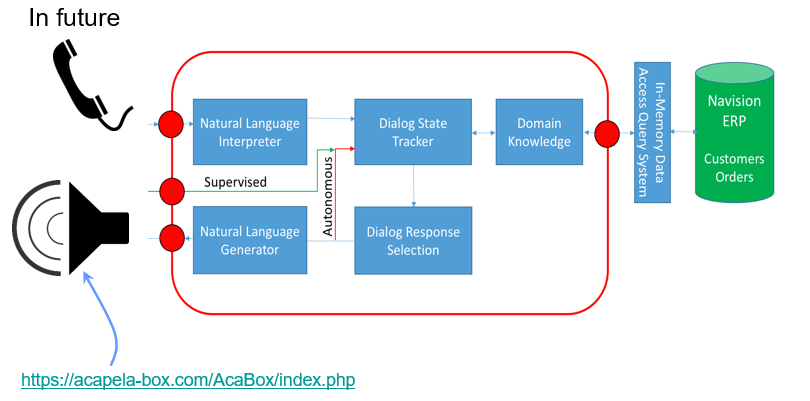


## Typical NLP Pipeline



## Example: Email & Chatbot Pipeline

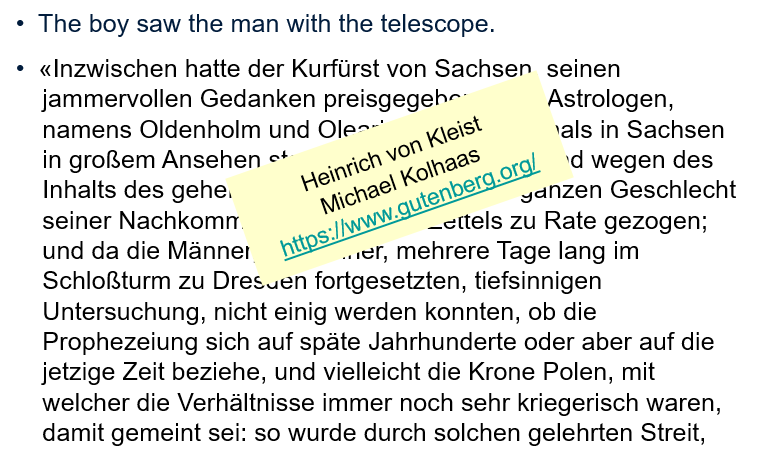




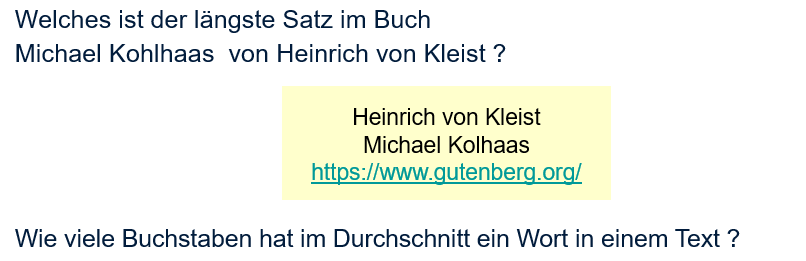
## Genres of Text

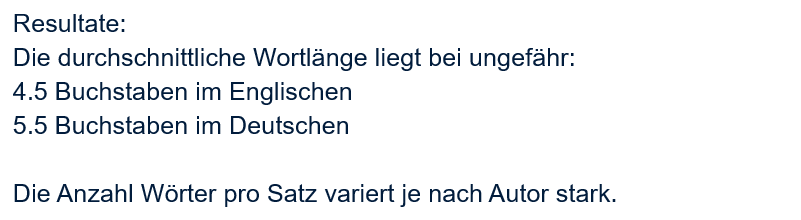


## Two Examples



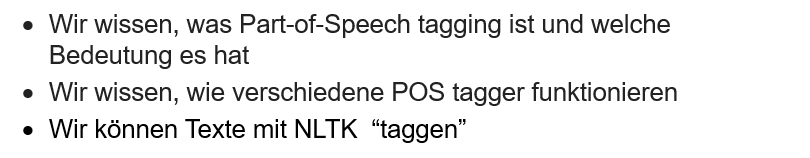
## Praktischer Teil



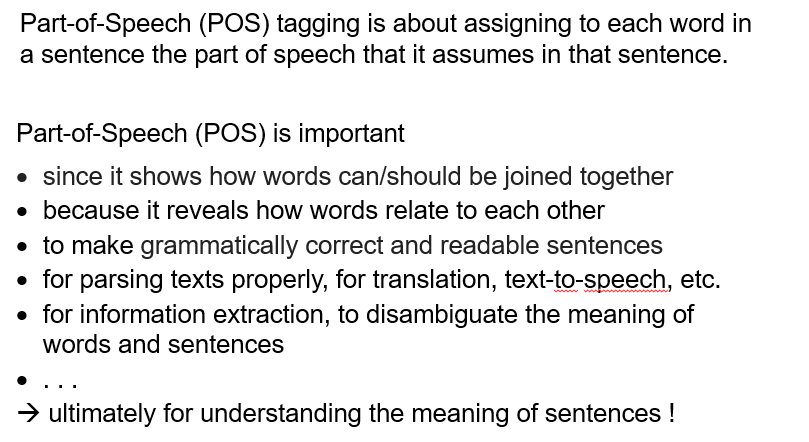


# Woche 2

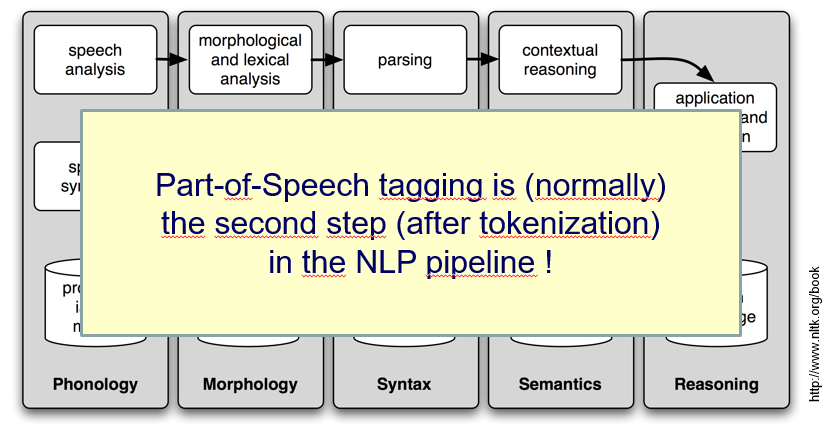
## Lernziele Woche 2



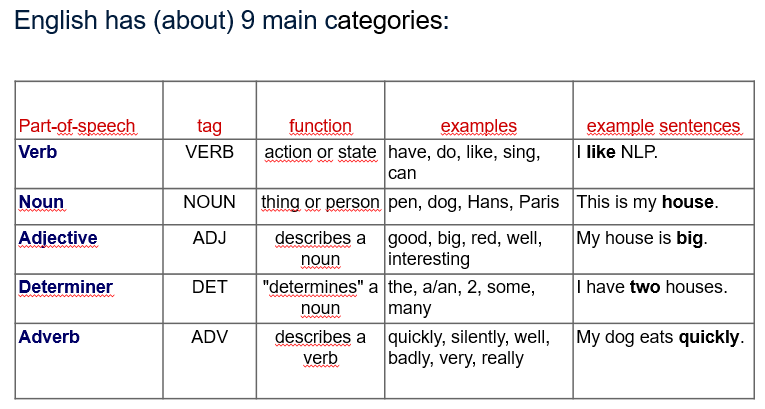
## Part-of-Speech

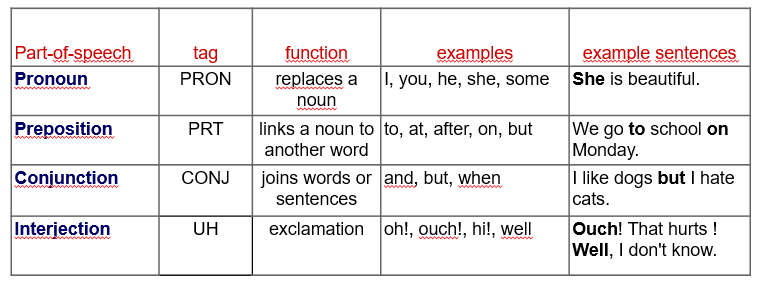


## Typical NLP Pipeline

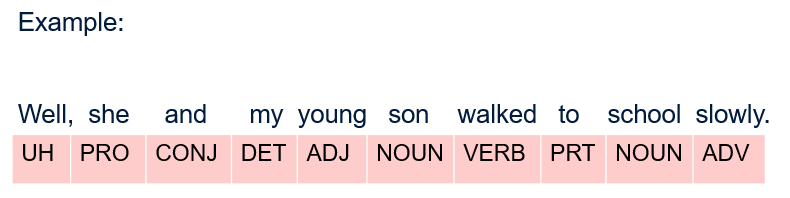


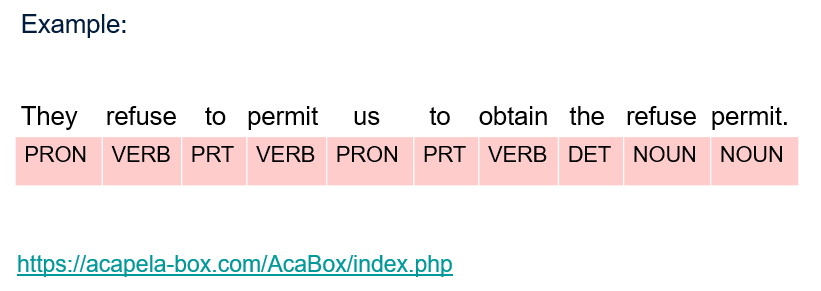
## Part-of-Speech

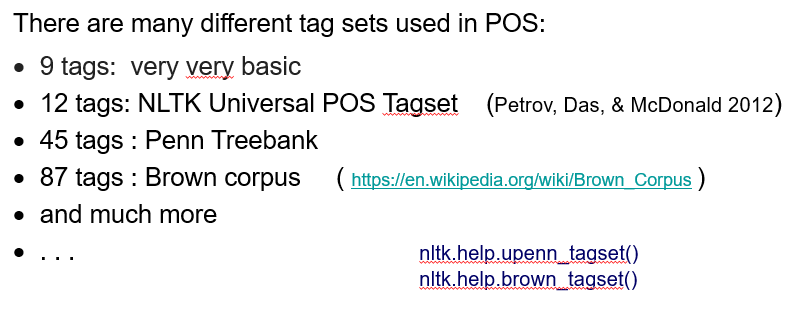


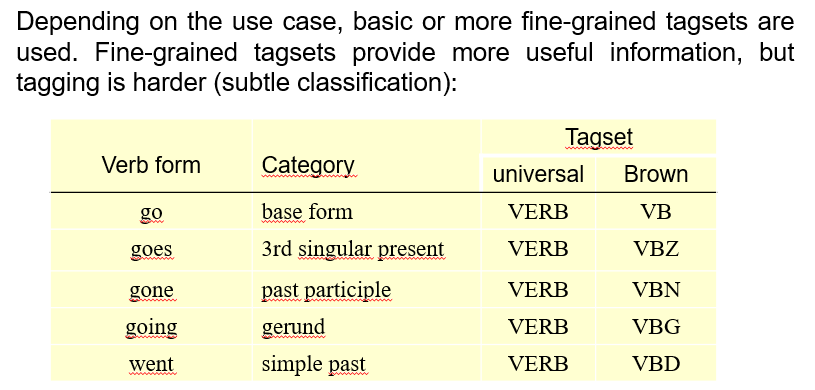


## Tagset

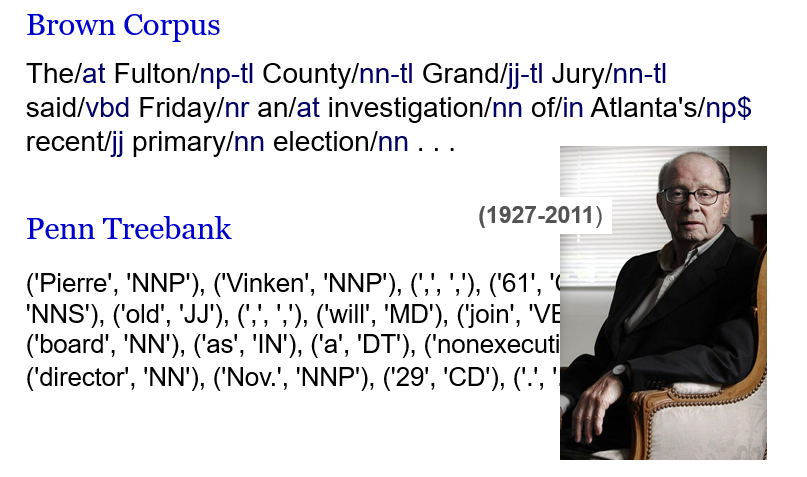






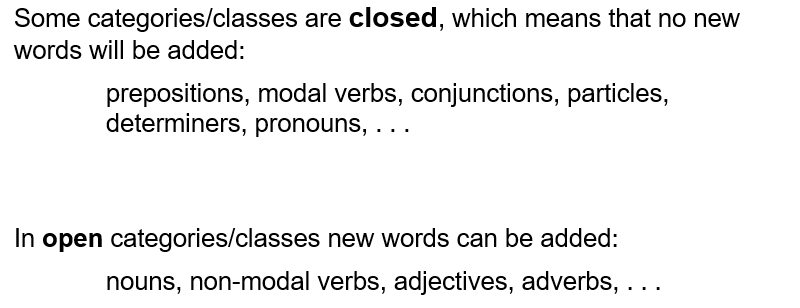


## Tagged Corpora

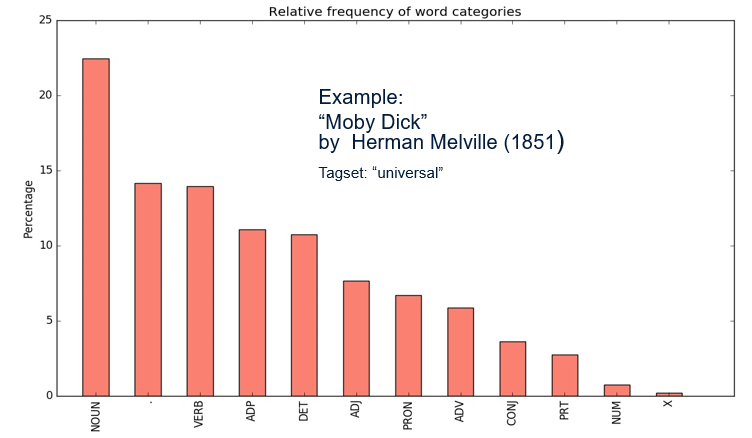


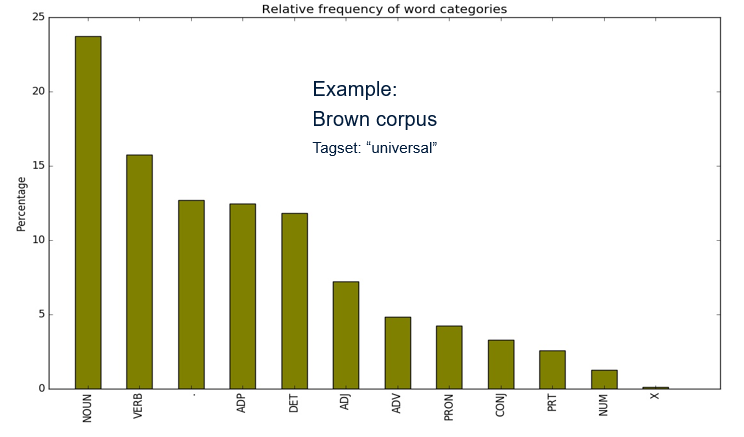


## Part-of-Speech

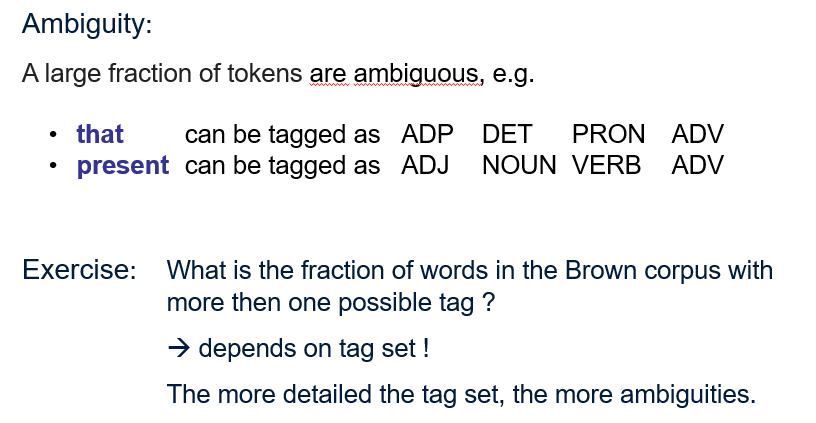


## Part-of-Speech distribution





## The POS problem



# Woche 3