Dependency Parsing

Kilian Evang, Jakub Waszczuk

Heinrich Heine Universität

Summer semester 2021

How does it work?

- Language: English and German

- Lectures: Tuesday, 14:30-16:00, online

- Labs: Friday, 10:30-12:00, online

- Weekly homework exercises (in groups)

- BN: homework

- AP: homework + exam

Where to find more?

- Website:

  https://user.phil.hhu.de/~waszczuk/teaching/hhu-dp-su21/

What is it about?

- Mainly *data-driven* dependency parsing

- Blocks:
  - Basics, machine learning, grammar-based methods
  - Transition-based methods
  - Implementation of a transition-based parser
  - Graph-based methods

- Supports used:
  - Book: Dependency parsing, by Sandra Kübler, Ryan McDonald and Joakim Nivre (https://katalog.ulb.hhu.de/Record/003823743)
  - Slides: used by the same authors for various courses (ACL 2006, ESSLLI 2007, EACL 2014)

Motivations

- Why syntax?

- Why parsing?

- Why is parsing hard/interesting?

Why syntax?

1. Grammaticality
2. Structure

Grammaticality:

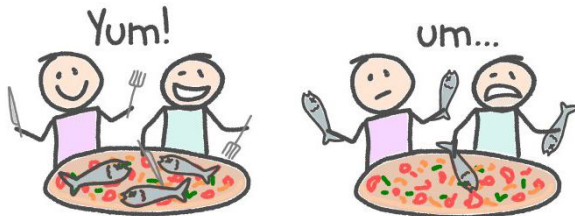The syntax of a language governs which combinations of words form sentences.

- The man walks
- *Walks man the
- #Man the walks

Why syntax?

> **Structure:**
>
> Syntax assigns structure to sentences.

"They ate the pizza with anchovies"

Creative Commons Attribution-NonCommercial 2.5
James Constable, 2010

Why parsing?

Academic interest: i.e.,

- Does this theory actually work?
- Can it be implemented?

Practical applications: e.g.,

- Question answering (e.g., IBM Watson – McCord et al., 2012)
- Machine translation
- etc.

Why is parsing hard?

Ambiguity

Ambiguity is the key problem in NLP.

- "I see a bird"

Every word in this sentence is ambiguous! (Church, ANLP 1988)

Why is parsing hard?

Ambiguity

Ambiguity is the key problem in NLP.

- "I see a bird"

Every word in this sentence is ambiguous! (Church, ANLP 1988)

- I: roman numeral?
- see: Holy see?
- a: French preposition?
- bird: verb?

Most of this ambiguity invisible to humans,
But computers need to be told how to resolve it!

Why is (dependency) parsing interesting?

- Universal Dependencies (http://universaldependencies.org)
  - Many languages
  - Many sentences
  - Each annotated in same way

  $\implies$ Lots of data for testing new parsing methods ("Common Task Framework")
  $\implies$ Cross-language studies, model transfer
  $\implies$ Nice balance between theory and practice