Introduction to Computational Linguistics
Session 1: Introduction and Definition

Dr. Stephen Bodnar, Denise Löfflad

Universität Tübingen

October 25, 2023

Plan

1 Course Organization

2 Definition

3 Different Perspectives on the Field

4 Different Methodologies

5 Relation to General Linguistics

6 References

Organization

See syllabus
```
https://moodle.zdv.uni-tuebingen.de/pluginfile.php/34474/
mod_resource/content/1/syllabus.pdf
```

Organization – Exam

- the exam will be held separately from the Java 1 exam
- the grades of Java 1 and Intro to CL will be combined (2/3 Java, 1/3 Intro to CL) for the grade of the module exam ("Modulklausur")

Organization

Please let us know about your background:

Introduction

What is Computational Linguistics?

What do you imagine Computational Linguistics is about ?

- definition
- characteristics
- applications

Definition

Computational Linguistics

Definition

Computational Linguistics

- bringing Language and Computers together

Definition - Association for Computational Linguistics (ACL)

"Computational linguistics is the scientific study of language from a computational perspective. Computational linguists are interested in providing computational models of various kinds of linguistic phenomena. These models may be "knowledge-based" ("hand-crafted") or "data-driven" ("statistical" or "empirical")."

`https://www.aclweb.org/portal/what-is-cl`

Definition - Jurafsky & Martin

"interdisciplinary field with many names corresponding to its many facets, names like speech and language processing, human language technology, natural language processing, computational linguistics, and speech recognition and synthesis. " [Jurafsky and Martin, 2009, p. 1]

Goal for CL

" The goal of this field is to get computers to perform useful tasks
involving human language " [Jurafsky and Martin, 2009, p. 1]

Natural language vs. formal language

"In contrast to artificial languages such as programming languages and mathematical notations, natural languages have evolved as they pass from generation to generation, and are hard to pin down with explicit rules" [Bird et al., 2009, p. 9]

Definition - Grishman

"Computational Linguistics is the study of computer systems for understanding and generating natural language." [Grishman, 1986, p. 4]

Different Perspectives on the Field

- theory perspective: describing/defining content and format of descriptions of phenomena
- method perspective: developing tools, procedures, formalism
- task perspective: applying methods to solve a specific task

Different Methodologies

- rule-based approaches: explicitly model language
- statistical approaches: implicitly model language

How does language relate to this ?

Relation to General Linguistics

"General Linguistics is concerned with human language as a universal and recognizable part of human behavior and of the human faculties, perhaps one of the most essential to human life as we know it, and one of the most far-reaching of human capabilities in relation to the whole span of mankind's achievements[. . .]Language in all its forms and manifestations, that is all the languages of the world and all the different uses to which in the various circumstances of mankind they are put, constitutes the field of the linguist." [Robins, 2014, p.2-3]

Relation to General Linguistics

" [C]omputational and theoretical linguists have rather different approaches and outlook. Computational linguists have been concerned with developing procedures for handling a useful range of natural language input…Theoretical linguists, in contrast, … are concerned with language universals - principles of grammar which apply to all natural languages " [Grishman, 1986, p. 6]

Natural Language Processing

There are two main subtasks of **Natural Language Processing (NLP)**:

- **Natural Language Understanding (NLU)**
- **Natural Language Generation (NLG)**

Natural Language Understanding (NLU)

1. Mapping the given input in natural language into useful representations, which involves analyzing different aspects of the input:
   - lexicon (e.g. **POS tagging**)
   - morphology (e.g. **lemmatization**)
   - syntax (e.g. **dependency parsing**)
   - discourse (e.g. **anaphora resolution**)
   - pragmatics (e.g. **sentiment detection**)

Natural Language Generation (NLG)

2. Producing meaningful natural language text from some more abstract representation
   - difficult to get right, very subtle problems
   - much less developed than NLU
   - very domain-specific (generating weather reports $\neq$ writing poems)

Applications of NLU and NLG

- Where can you find these two technologies being used today?
- Where might we see them in the future?

Next class

- No lecture next Wednesday, Nov. 1st (public holiday)
- Instead, Reading Assignment:
  Chapter 1 in Dickinson et al. [2012] "Language and Computers"

References and Acknowledgments

The contents about the historical perspective are adapted from Frank Richter's *Introduction to Computational Linguistics* seminar from the winter semester 2012/13.

Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.", 2009.

Markus Dickinson, Chris Brew, and Detmar Meurers. *Language and computers*. John Wiley & Sons, 2012.

Ralph Grishman. *Computational linguistics: an introduction*. Cambridge University Press, 1986.

Daniel Jurafsky and James H Martin. Naïve bayes classifier approach to word sense disambiguation. *Computational Lexical Semantics*, 2009.

Robert Henry Robins. *General linguistics*. Routledge, 2014.