Lecture 1: Introduction

Computational Linguistics

John Gamboa

General things about this lecture

About the Lecture in General

About the Lecture in General

- Relatively "new" lecture (this is its third iteration)

About the Lecture in General

- Relatively "new" lecture (this is its third iteration)

- Primary audience: Cognitive Science students
  - Will cover some basics about AI and programming
  - (Comparatively) light on maths and technical details

About the Lecture in General

- Relatively "new" lecture (this is its third iteration)

- Primary audience: Cognitive Science students
  - Will cover some basics about AI and programming
  - (Comparatively) light on maths and technical details

- What is important:
  - See the big picture (roadmap for going deeper)
  - Understanding the basic ideas

About the Lecture in General

Two more points:

About the Lecture in General

Two more points:

1. Ask **questions**!
   - I need feedback
   - This makes the course more interesting
   - Questions help you to understand

About the Lecture in General

Two more points:

1. Ask **questions**!
   - I need feedback
   - This makes the course more interesting
   - Questions help you to understand

2. Ask *critical* **questions**!

Topics

- Jupyter notebooks, Python basics, NLTK
- Accessing text corpora
- Corpus linguistics
- Tokenization, regular expressions, filtering
- Machine learning basics
- Artificial neural networks
- POS tagging, stemming
- WordNet
- Distributional semantics
- Some signal processing

Topics

- Jupyter notebooks, Python basics, NLTK
- Accessing text corpora
- Corpus linguistics
- Tokenization, regular expressions, filtering
- Machine learning basics
- Artificial neural networks
- POS tagging, stemming
- WordNet   (still subject to a few changes)
- Distributional semantics
- Some signal processing

Topics

Not so much / not at all:

- Formal linguistics (e.g. formal semantics, types of grammars etc.)
- Specific languages
- State-of-the-art for individual applications

Topics

Not so much / not at all:

- Formal linguistics (e.g. formal semantics, types of grammars etc.)
- Specific languages
- ~~State of the art for individual applications~~
    (this year I'd like to try out speaking about the more gore stuff)

Organization

Organization

- Credits: 3 or 4 CP (?)

Organization

- Credits: 3 or 4 CP (?)

- Questionnaires every week (this is your "attendance")

Organization

- Credits: 3 or 4 CP (?)

- Questionnaires every week (this is your "attendance")

- Formal requirements:
    - Exam (i.e., score 65% of the exam), only for non-Cognitive Sciences students

Time commitment for this course

- Credits: 3 or 4 CP (?)

Time commitment for this course

- Credits: 3 or 4 CP (?)

- 3 to 6 hours of work per week
    - 6h / week for 14 weeks = 84h
    - 1 CP = 28h → 3 CP = 84h

Technological Requirements

You'll need…

Technological Requirements

You'll need…

- a computer / cellphone for the lectures;

Technological Requirements

You'll need…

- a computer / cellphone for the lectures;

- you should be able to run Python

Technological Requirements

You'll need…

- a computer / cellphone for the lectures;

- you should be able to run Python

- you should be able to access the synchronous classes

Technological Requirements

You'll need…

- a computer / cellphone for the lectures;

- you should be able to run Python

- you should be able to access the synchronous classes

- *Additionally:* **no fear from your computer xP**

Behavioral Guidelines

- You are encouraged to post anything in the forum

Behavioral Guidelines

- You are encouraged to post anything in the forum

- Still, please avoid...
  - trolling,
  - bullying, or
  - cursing

Today

- Introduction (DONE)
- General things about this lecture (DONE)
  - About this lecture
  - Topics
  - Miscellaneous organizational points
- What is computational linguistics?
  - Areas
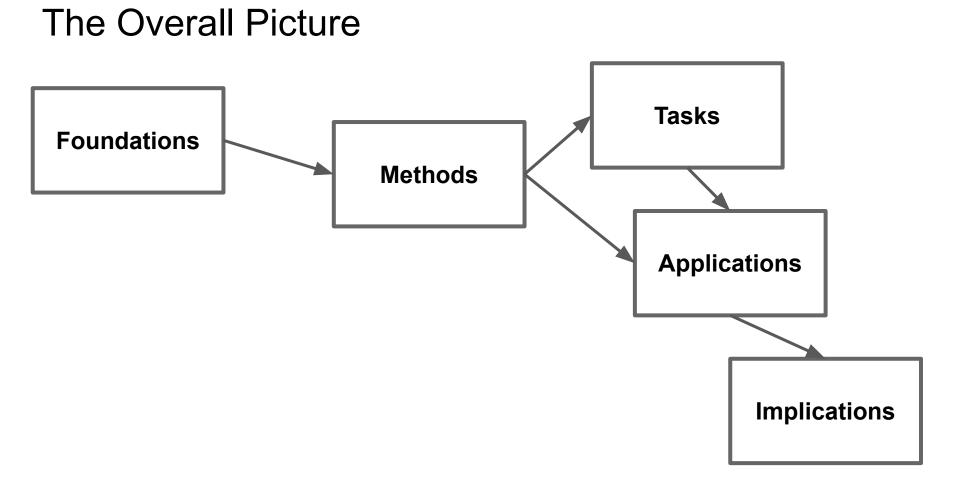 - Relations to other fields
  - Methods
  - Tasks
  - Applications
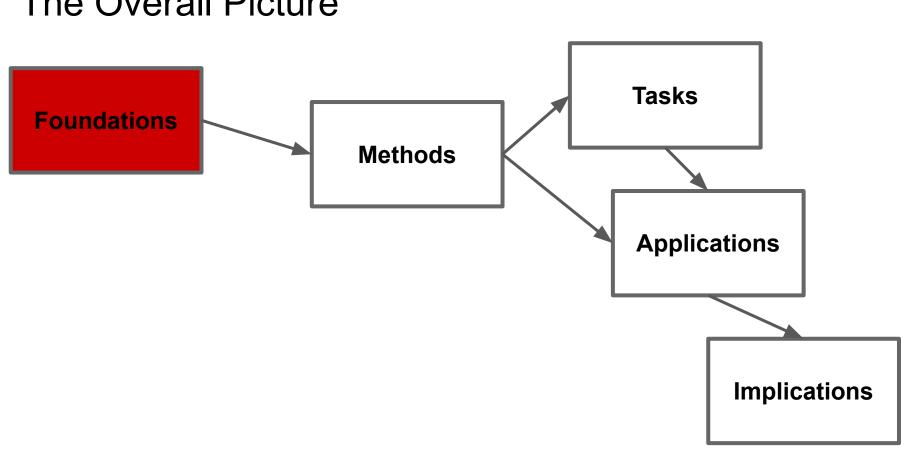 - Implications

What is Computational Linguistics?

What is Computational Linguistics?

A working definition:

"*Computational linguistics* is an <u>interdisciplinary</u> field concerned with the <u>statistical or rule-based modeling</u> of <u>natural language</u> from a computational perspective, as well as the study of appropriate <u>computational approaches to linguistic questions</u>."

(Wikipedia page on "computational linguistics"; 2018)
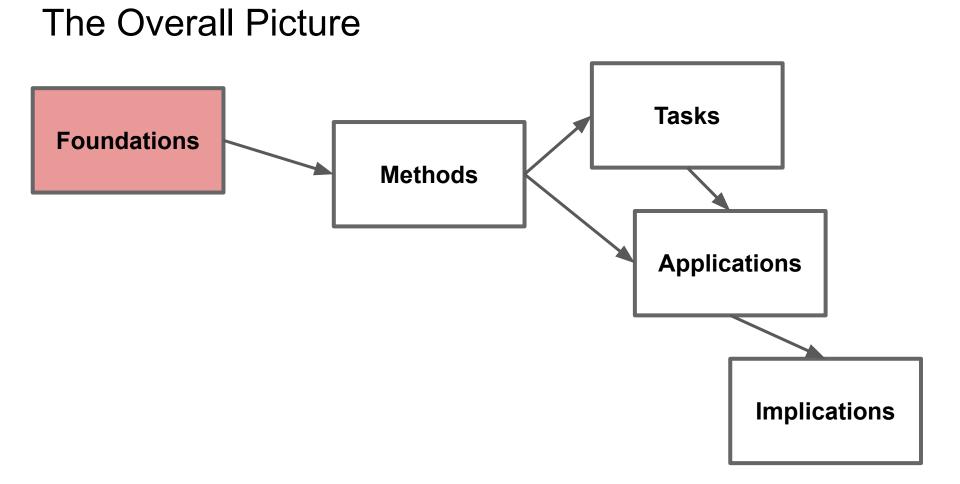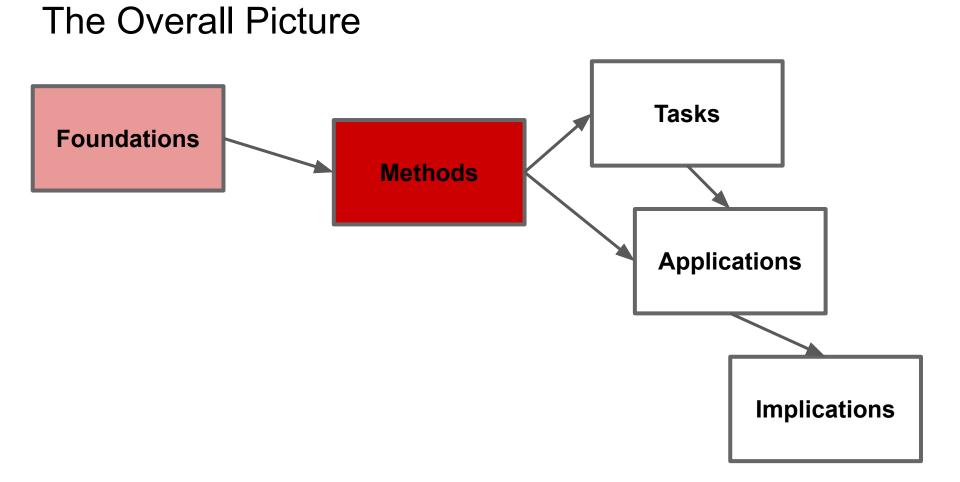
The Overall Picture

The Overall Picture

The Overall Picture

Foundations

Foundations
- Linguistics (syntax, semantics, pragmatics, …)
- Computer science (general data processing, machine learning, interfaces, …)
- Mathematics (statistics, logic, foundation of models)
- Philosophy (e.g. philosophy of meaning)
- Psychology (e.g. emotion processing)

Implications

The Overall Picture

The Overall Picture

Methods

- Procedural
  - Pattern-based approaches (e.g. ELIZA)
  - Rule-based approaches (e.g. SHRDLU)

Methods

- Procedural
  - Pattern-based approach
  - **Rule-based** approaches

Conversation with SHRDLU (Example)
Person: Pick up a big red block.
Computer: OK.
Person: Grasp the pyramid.
Computer: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.
Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.
Computer: BY "IT", I ASSUME YOU MEAN THE BLOCK WHICH IS TALLER THAN THE ONE I AM HOLDING.
Computer: OK.
Person: What does the box contain?
Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.
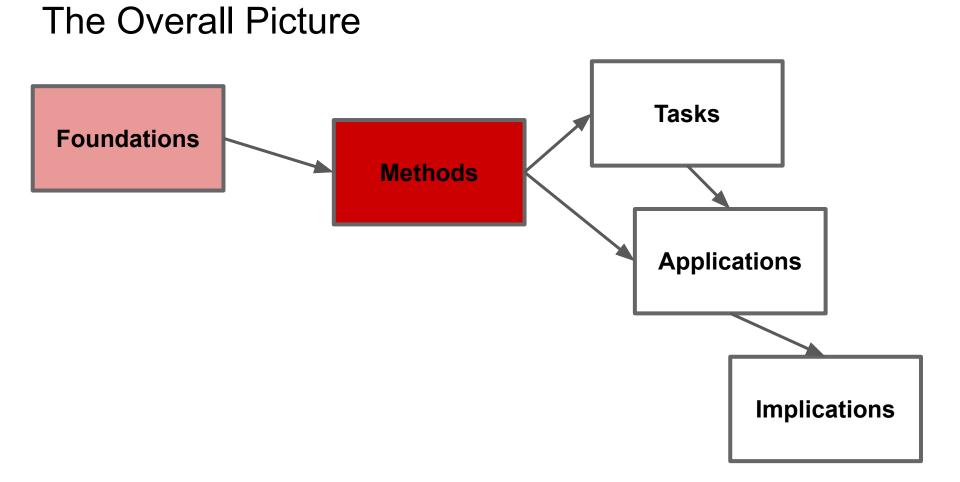Person: What is the pyramid supported by?
Computer: THE BOX.
…

Methods

- Procedural
  - Pattern-based approaches (e.g. ELIZA)
  - Rule-based approaches (e.g. SHRDLU)
- Structure- and knowledge-based
  - Generalized Phrase Structure Grammar (GPSG), Head-driven Phrase Structure Grammar (HPSG), …
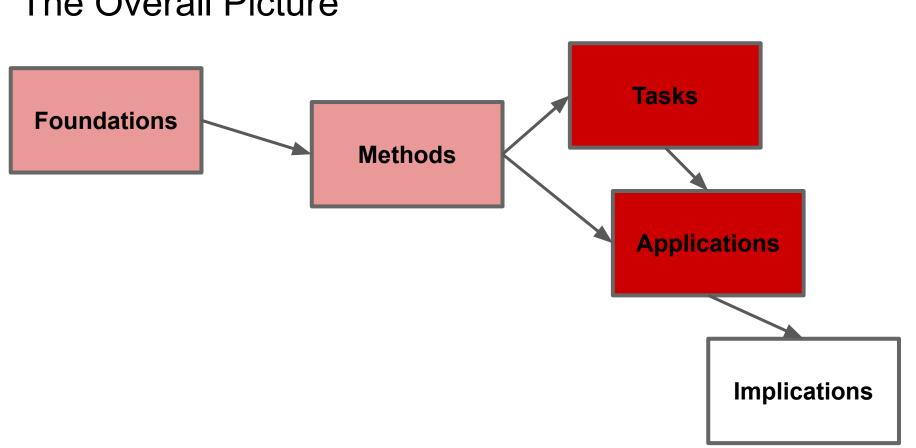
Methods

- Procedural
  - Pattern-based approaches (e.g. ELIZA)
  - Rule-based approaches (e.g. SHRDLU)
- Structure- and knowledge-based
  - Generalized Phrase Structure Grammar (GPSG), Head-driven Phrase Structure Grammar (HPSG), …
- Corpus-based, statistical
  - E.g. support vector machines, artificial neural networks, ...

Methods

- Procedural
  - Pattern-based approaches (e.g. ELIZA)
  - Rule-based approaches (e.g. SHRDLU)
- Structure- and knowledge-based
  - Generalized Phrase Structure Grammar (GPSG), Head-driven Phrase Structure Grammar (HPSG), …
- Corpus-based, statistical
  - E.g. support vector machines, artificial neural networks, ...

Trend: Combine statistical with symbolic approaches

The Overall Picture

The Overall Picture

Applications

Tasks

- Spell and grammar checking

Applications

- Spell and grammar checking

Tasks

- Parsing language
- POS tagging

- …

Applications

- Spell and grammar checking
- Machine translation

Tasks

- Language generation
- Parsing language
- POS tagging

- Representing information

- Word sense disambiguation
- Dataset creation
- …

Applications

- Spell and grammar checking
- Machine translation
- Document retrieval and clustering applications

Tasks

- Language generation
- Parsing language
- POS tagging

- Information extraction
- Representing information

- Creating statistics
- Text classification
- Word sense disambiguation
- Dataset creation
- …

Applications

- Spell and grammar checking
- Machine translation
- Document retrieval and clustering applications
- Information extraction and summarization

Tasks

- Language generation
- Parsing language
- POS tagging
- Emotion recognition
- Sentiment analysis
- Information extraction
- Representing information
- Summarization

- Creating statistics
- Text classification
- Word sense disambiguation
- Dataset creation
- …

Applications

- Spell and grammar checking
- Machine translation
- Document retrieval and clustering applications
- Information extraction and summarization
- Chatbots and companionable dialogue agents

Tasks

- Dialogue modeling and planning
- Language generation
- Parsing language
- POS tagging
- Emotion recognition
- Sentiment analysis
- Information extraction
- Representing information
- Summarization
- Speech recognition
- Speech synthesis
- Creating statistics
- Text classification
- Word sense disambiguation
- Dataset creation
- …

Applications

- Spell and grammar checking
- Machine translation
- Document retrieval and clustering applications
- Information extraction and summarization
- Chatbots and companionable dialogue agents
- Virtual worlds, games, and interactive fiction

Tasks

- Dialogue modeling and planning
- Language generation
- Parsing language
- POS tagging
- Emotion recognition
- Sentiment analysis
- Information extraction
- Representing information
- Summarization
- Speech recognition
- Speech synthesis
- Creating statistics
- Text classification
- Word sense disambiguation
- Dataset creation
- …

Applications

- Spell and grammar checking
- Machine translation
- Document retrieval and clustering applications
- Information extraction and summarization
- Chatbots and companionable dialogue agents
- Virtual worlds, games, and interactive fiction
- Natural language user interfaces
  - Question answering (text-based or knowledge-based)
  - Database front-ends
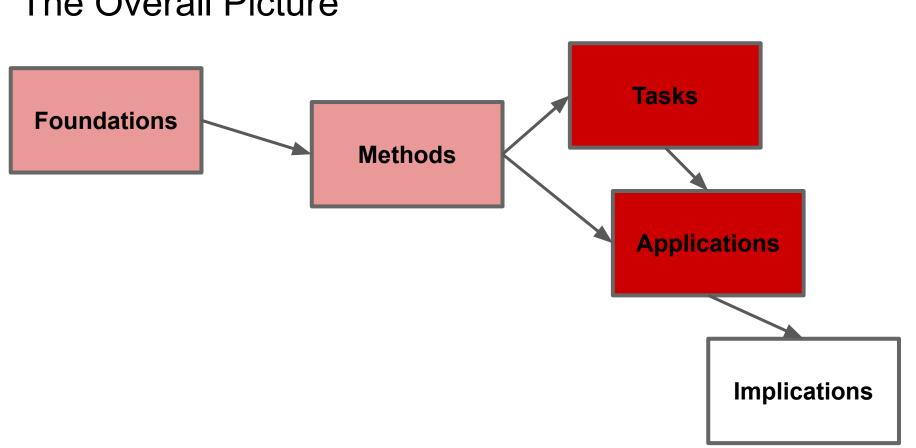 - Voice-based web services and assistants

Tasks

- Dialogue modeling and planning
- Language generation
- Parsing language
- POS tagging
- Emotion recognition
- Sentiment analysis
- Information extraction
- Representing information
- Summarization
- Speech recognition
- Speech synthesis
- Creating statistics
- Text classification
- Word sense disambiguation
- Dataset creation
- …

Applications

- Spell and grammar checking
- Machine translation
- Document retrieval and clustering applications
- Information extraction and summarization
- Chatbots and companionable dialogue agents
- Virtual worlds, games, and interactive fiction
- Natural language user interfaces
  - Question answering (text-based or knowledge-based)
  - Database front-ends
  - Voice-based web services and assistants
- Collaborative problem solvers and intelligent tutors
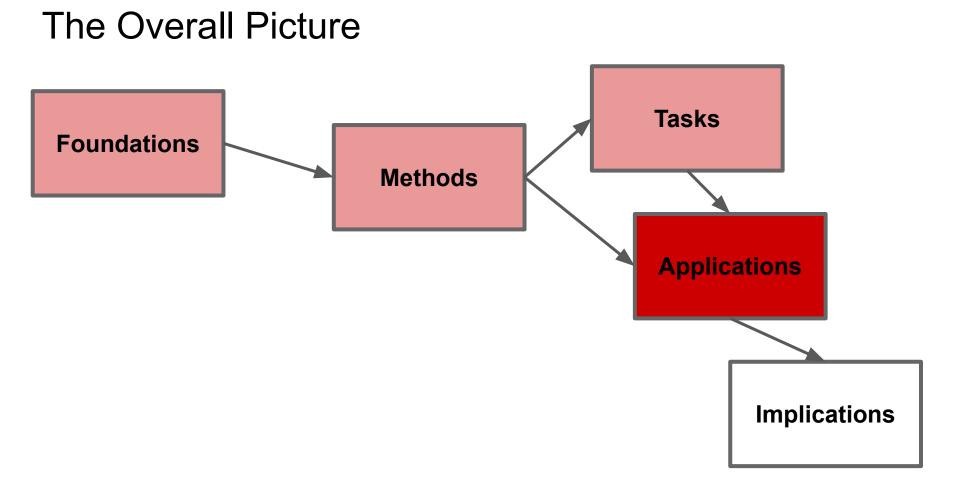
Tasks

- Dialogue modeling and planning
- Language generation
- Parsing language
- POS tagging
- Emotion recognition
- Sentiment analysis
- Information extraction
- Representing information
- Summarization
- Speech recognition
- Speech synthesis
- Creating statistics
- Text classification
- Word sense disambiguation
- Dataset creation
- …

Applications

- Spell and grammar checking
- Machine translation
- Document retrieval and clustering applications
- Information extraction and summarization
- Chatbots and companionable dialogue agents
- Virtual worlds, games, and interactive fiction
- Natural language user interfaces
  - Question answering (text-based or knowledge-based)
  - Database front-ends
  - Voice-based web services and assistants
- Collaborative problem solvers and intelligent tutors
- Language-enabled robots

Tasks

- Dialogue modeling and planning
- Language generation
- Parsing language
- POS tagging
- Emotion recognition
- Sentiment analysis
- Information extraction
- Representing information
- Summarization
- Speech recognition
- Speech synthesis
- Creating statistics
- Text classification
- Word sense disambiguation
- Dataset creation
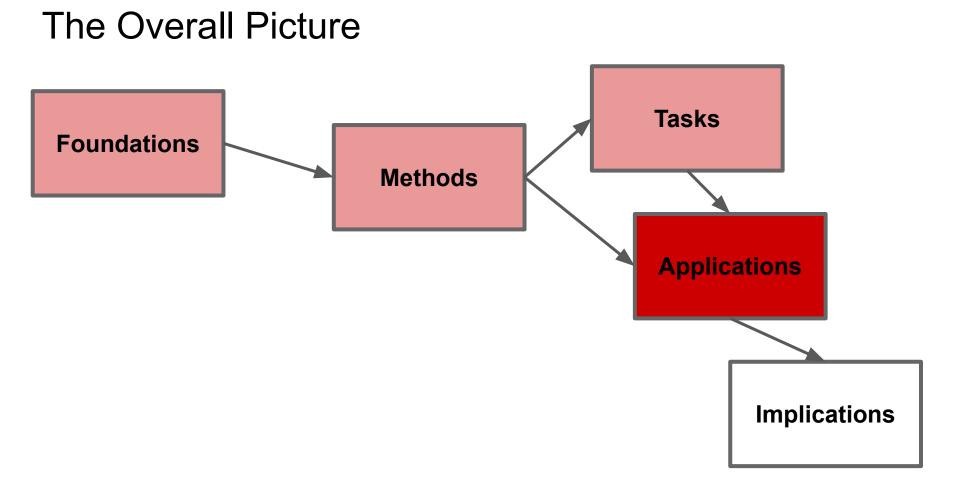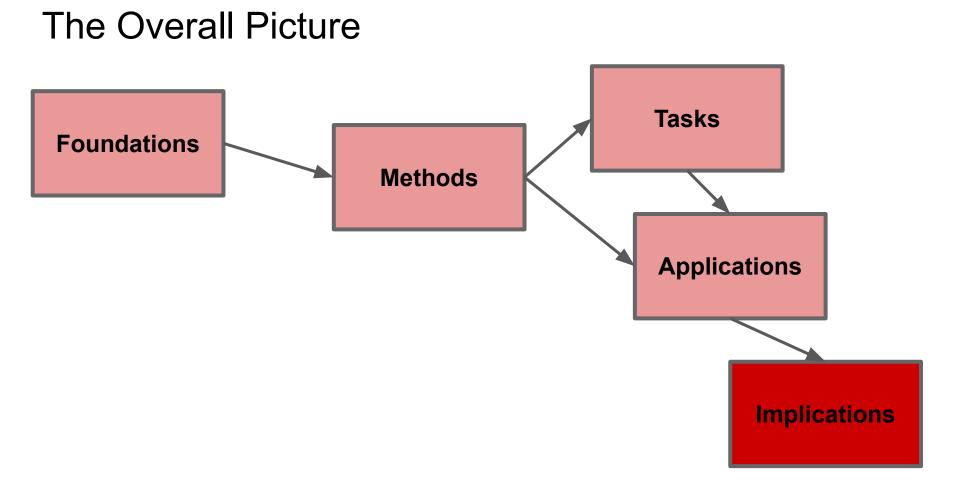- …

The Overall Picture

The Overall Picture

Applications in research

- Quantitative testing of theories (e.g. language acquisition, historical development)

- Analyzing trained models to develop hypotheses

- Analyzing large-scale correlations that are psychologically interesting (e.g. estimating emotional state of population from social media posts)

- ...

The Overall Picture

The Overall Picture

Implications

Many of these applications can make our lives more convenient.

Implications

Many of these applications can make our lives more convenient.

But what would be potential risks?
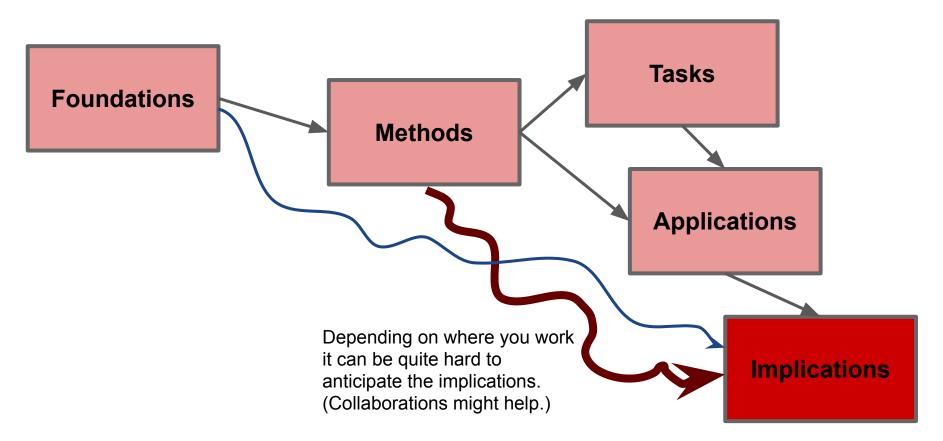Anything problematic about what we've seen so far?

Implications

Many of these applications can make our lives more convenient.

But what would be potential risks?
Anything problematic about what we've seen so far?

- Monitoring emotions can give you control
- Chatbots are also used for propaganda
- Data and privacy concerns ("Big data"?!)

The Overall Picture

Foundations → Methods → Tasks → Applications → Implications

Depending on where you work it can be quite hard to anticipate the implications. (Collaborations might help.)

Discussion

Computational linguistics might be important...

But is it also *ethically "good"*?

Discussion

Computational linguistics might be important...

But is it also *ethically "good"*?

What do *you* want to use it for?

Today

- Introduction (DONE)
- General things about this lecture (DONE)
  - About this lecture
  - Topics
  - Miscellaneous organizational points
- What is computational linguistics? (DONE)
  - Areas
  - Relations to other fields
  - Methods
  - Tasks
  - Applications
  - Implications

Reading

Today:

- https://plato.stanford.edu/entries/computational-linguistics/
- https://en.wikipedia.org/wiki/Computational_linguistics

This course in general:

- "Computational Linguistics" by Bolshakov and Gelbukh:
  https://www.gelbukh.com/clbook/
- "Natural language processing with Python" by Bird, Klein, Loper:
  http://www.nltk.org/book/
- Sam Bowman's Introduction slides to Computational Linguistics