Introduction to natural language processing

Herman Kamper

What is natural language processing?

Course information

Course goals and philosophy

Google assistant

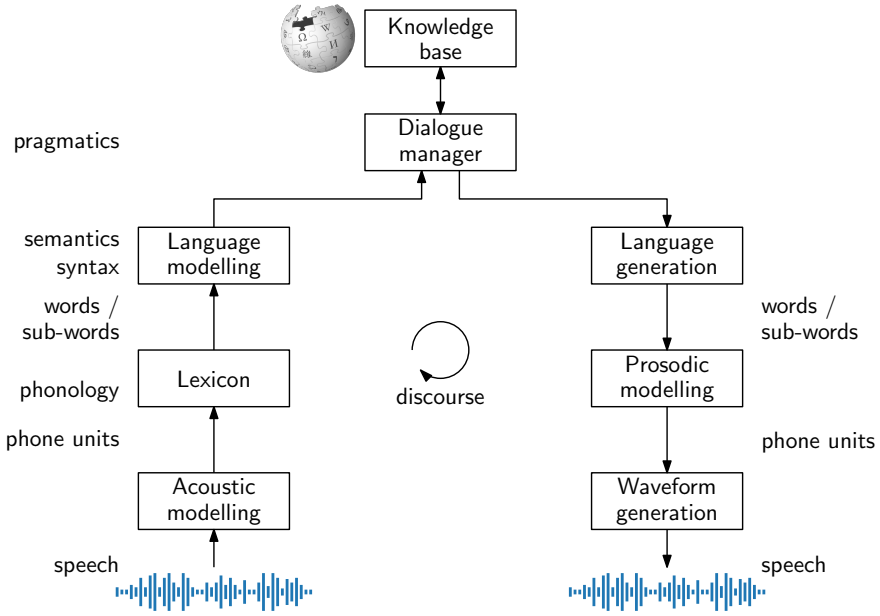
```
User:   Okay Google, what's the weather like in
        Stellenbosch today?

Google: Today's forecast for Stellenbosch is
        twenty-two degrees and sunny.

User:   What about tomorrow?

Google: Tomorrow's forecast for Stellenbosch is
        nineteen degrees and partly cloudy.
```

Think about all the components needed in the system for this brief conversation:[1]



[1] Figure adapted from https://zerospeech.com/.

What is natural language processing?

Natural language processing (NLP) aims to enable computers to process human language in order to perform useful tasks.

Computational linguistics uses computers to discover and better understand the principles of human language. In practice, the term is often synonymous with NLP (as is evident in the names of the big NLP conferences). But there is a somewhat more scientific rather than a task focus.

Spoken language processing deals specifically with continuous speech signals. This typically involves either converting a speech waveform into categorical units (recognition) or transforming these units back into a waveform (generation or synthesis). This corresponds to the lower part of the figure above.

All these areas overlap. NLP often refers specifically to processing symbolic text input: the upper part of the figure.

Examples of NLP applications

- Spam detection
- Text classification: Grammarly's tone detection predicting whether text is friendly or formal
- Machine translation: Google Translate
- Autocomplete and smart compose: Gmail
- Virtual assistants: Siri, Cortana, Google Assistant
- General dialogue systems: ChatGPT
- Coding assistants: Copilot, CodeT5

Note on ChatGPT

ChatGPT combines several of the modules at the top of the figure above: Dialogue management, queries to a knowledge base and language generation are all solved using a next-word-predicting language model. The knowledge base can be an internet search engine or a Python interpreter.

Although GPT combines the different sub-modules into one, it is still useful to understand the separate tasks. This could help us better understand how GPT accomplishes these amazing feats. It will also help us assess its limitations.

NLP817 course information

Instructor

Mr. Reuben Smit (28583272@sun.ac.za)

Textbook

> D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed. draft, 2026.

A free draft is available online. I will refer to this as J&M3 in the notes. Older editions will be denoted as J&M1 and J&M2.

Flipped classroom

The course will be presented as a flipped classroom. Every week you will watch the videos for a set of units. You will then email me any questions that you have before the contact session. The contact session will consist of a section where I answer questions followed by a short test. The list of units will be posted weekly on STEMLearn.

Contact sessions

Venue: MLAI lab, 2nd floor of the General Engineering building

Dates and times: - Monday 30 March from 09:00 to 10:30 - Wednesday 1 April from 09:00 to 10:30 - Tuesday 7 April from 15:00 to 16:30 - Tuesday 14 April from 15:00 to 16:30 - Tuesday 21 April from 15:00 to 16:30

Assessments

- Class tests (35%)
- Assignment 1 (32.5%)
- Assignment 2 (32.5%)

For each assignment you will write a report in a paper format and upload your report and code on STEMLearn. The final cut-off on STEMLearn will be strictly enforced.

Course websites

- STEMLearn: https://stemlearn.sun.ac.za/course/view.php?id=3045
- Material: https://www.kamperh.com/nlp817/

NLP817 goals and philosophy

Goals of course

- Introduce basic tasks in NLP and see why these are challenging:
  - Outline the processing pipeline for a task
  - Datasets, models, algorithms and evaluation methods
- Introduce the algorithms and models used to solve these tasks:
  - Simulate algorithms step-by-step with pen and paper
  - Implement some algorithms and models in code
- Give enough background to be able to read current NLP research papers and do a research assignment in language or speech processing

Philosophical stuff that I hope you will take away

- Start simple, e.g. don't start with a VAE with a fancy prior
- Often we don't have a "correct answer"
- Basic machine learning methodology is crucial, e.g. data splits
- Writing is important

Course will be self-contained

Some of you might have extensive machine learning experience already. I will aim to make the course self-contained, which means I will explain several models from scratch. Even if you have seen these models before, hopefully these explanations in the context of a real problem will help you better understand the challenges in NLP, and maybe even help you understand the models themself better. Also, help those around you.

Why class tests?

- To force you to look at the material
- For me to see whether you are getting it
- For you to see whether you are getting it
- Testing has unfortunately been proven to be effective for learning
- You cannot use ChatGPT in a closed-book class test

Feedback

- If you feel the pace is too fast or slow, please let me know.
- If you spot *any* mistakes in the notes, please let me know. There are still tons of mistakes.
- If you love or hate the notes or lecturing style, please let me know. I want to get better.

Videos covered in this note

- What is natural language processing? (15 min)