ANLP Week 1 / Lecture 1
Introduction

Adam Lopez

Co-lecturers: Edoardo Ponti and Shay Cohen

(today's content based on slides by Sharon Goldwater)

School of informatics

Lecture recording

- **Lectures for this course are recorded.**
- We have also requested that they be streamed, but we haven't had time to test this!
- Recording normally works, but we can't guarantee it.
- The microphone picks up my voice, but not yours. (I will repeat questions/comments from students so they are recorded. Please remind me if I forget!)
- We can pause the recording / streaming at request. Signal to me if you want me to pause the recording at any time.
- Recordings will only be available to enroled students (you must log in).
- You may not re-distribute recordings or provide access to anyone outside the course.

Course Content

What is Natural Language Processing?

What is Natural Language Processing?

Applications

- Machine Translation
- Information Retrieval
- Question Answering
- Dialogue Systems
- Information Extraction
- Summarization
- Sentiment Analysis
- ...

Core problems

- Morphological analysis
- Syntactic analysis
- Named-entity recognition
- Coreference resolution
- Word sense disambiguation
- ...

This Course

Linguistics

- words
- morphology
- parts of speech
- syntax
- semantics
- (discourse?)

Computational methods

- finite state machines (morphological analysis, POS tagging)
- grammars and parsing (CKY, statistical parsing)
- probabilistic models and machine learning (HMMS, PCFGs, logistic regression, neural networks)
- vector spaces (distributional semantics)
- lambda calculus (compositional semantics)

What this course is, and what it isn't

This course is a fast-paced introduction/survey course. We **will**

- introduce many of the basic problems in NLP and discuss why they are challenging
- present basic linguistic concepts and computational methods (maths and algorithms) that can be used for these problems
- cover many conceptual building blocks of cutting edge NLP
- give you enough background to be able to read (some) NLP research papers and take follow-on courses in semester 2

But we **will not**

- cover cutting edge methods or heavy-duty machine learning. For these, you will need to take NLU+ in semester 2, and / or ML courses.

An Analogy: Joinery / Carpentry

Image credits: flikr users Gerard Stolk (top left); Dean Shareski (right)

An Analogy: Joinery / Carpentry

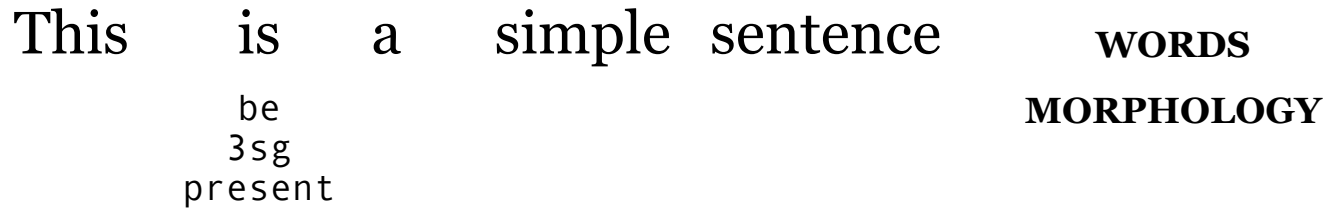
Why not ChatGPT / LLMs / Deep Learning??

- Toolkits for NLP and deep learning now allow you to easily put together basic NLP systems.
- But to go beyond basic recipes and make more effective systems for your specific needs, you must understand fundamental principles.
  - What are the linguistic problems and challenges?
  - What mathematical and modeling techniques are relevant?
  - Deep learning systems incorporate many concepts from basic models and algorithms we'll cover. Starting from first principles will help you understand them better when you see them.
- This course won't turn you into a master carpenter (that takes years!), but it will get you past basic recipes.

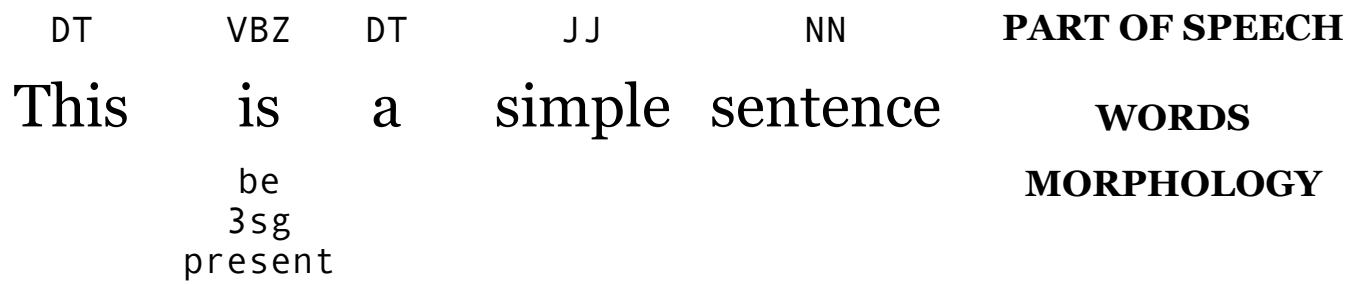
Words

This is a simple sentence **WORDS**

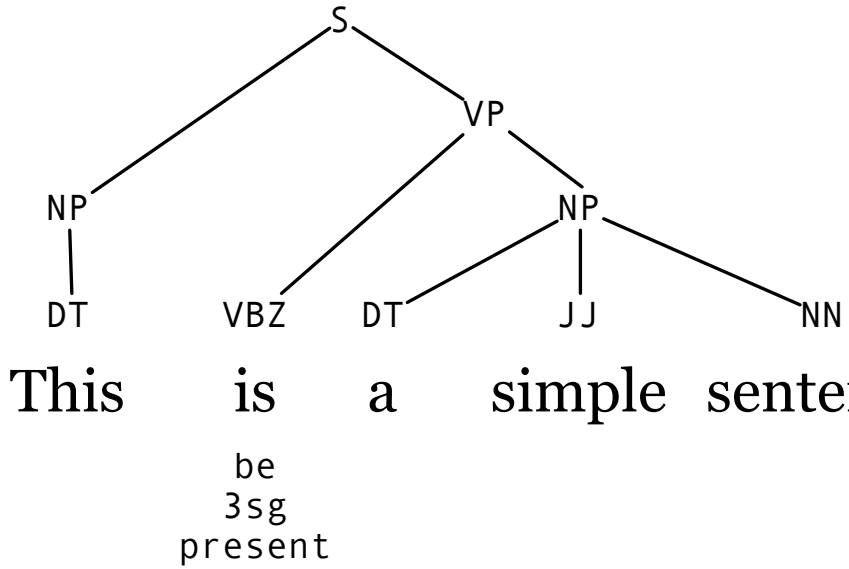
Morphology



Parts of Speech



Syntax



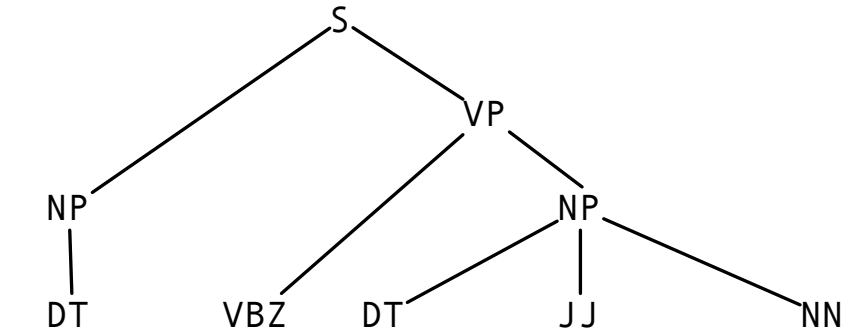
SYNTAX

PART OF SPEECH

WORDS

MORPHOLOGY

Semantics



be
3sg
present

SIMPLE1
having
few parts

SENTENCE1
string of words
satisfying the
grammatical rules
of a languauge

SYNTAX

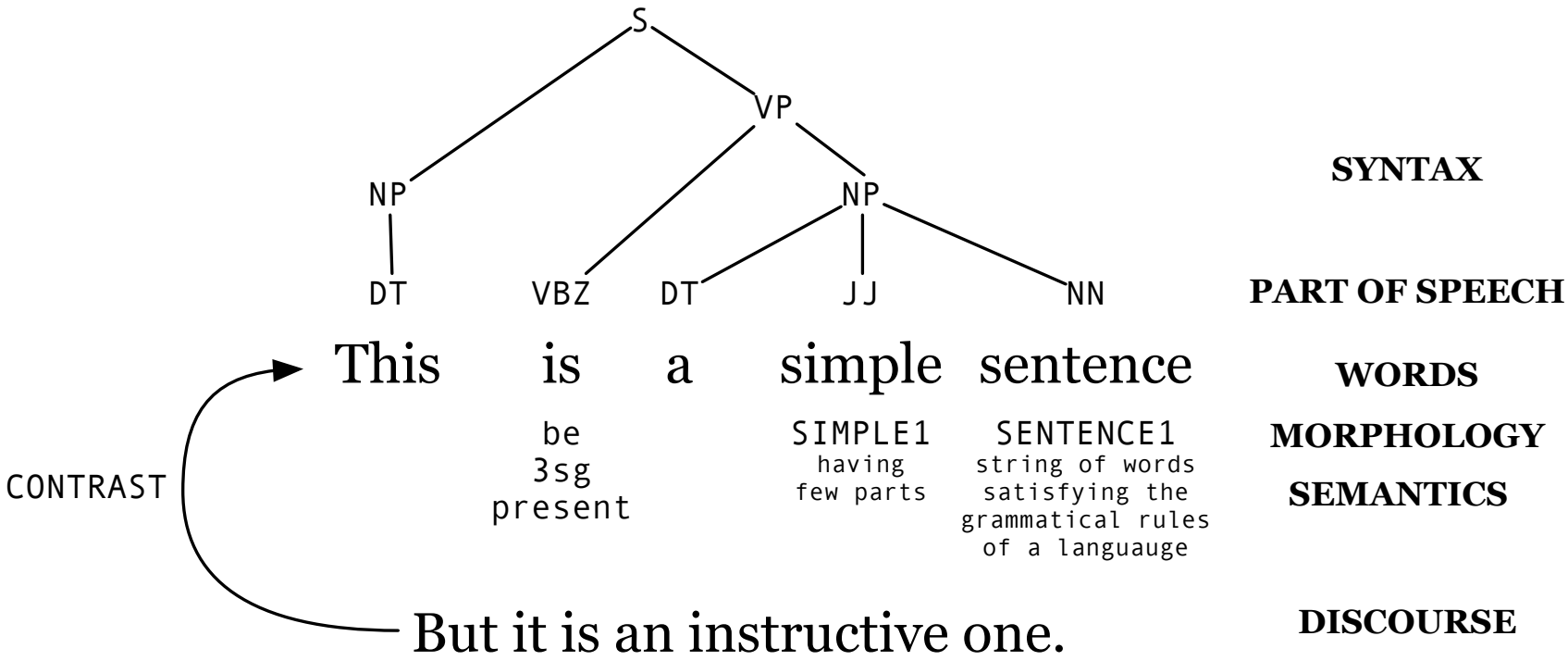
PART OF SPEECH

WORDS

MORPHOLOGY

SEMANTICS

Discourse


S
VP
NP
NP
DT
VBZ
DT
JJ
NN
This
is
a
simple
sentence
be
3sg
present
SIMPLE1
having
few parts
SENTENCE1
string of words
satisfying the
grammatical rules
of a languauge
CONTRAST
But it is an instructive one.
SYNTAX
PART OF SPEECH
WORDS
MORPHOLOGY
SEMANTICS
DISCOURSE

Why is Language Hard?

- Ambiguities on many levels, need context to disambiguate
- Rules, but many exceptions
- Language is infinite. We cannot see examples of everything, and the vast majority of what we do see is rare

Ambiguity

- Ambiguity is sometimes used intentionally for humor:

  I'm not a fan of the new pound coin, but then again, I hate all change.[1]

  One morning I shot an elephant in my pajamas. How he got in my pajamas I don't know.[2]

- What makes these jokes funny? Is it the same sort of ambiguity, or something different in each case?

[1] Ken Cheng, 2017. (Winner of Dave's Funniest Joke of the Fringe award.)
[2] Groucho Marx, in the 1930 film Animal Crackers.

Ambiguity

- However, ambiguity is much more common than jokes.
- Follow-up exercise: where is the ambiguity in these examples? Which is more like Joke 1? Joke 2?
  1. This morning I walked to the bank.
  2. I met the woman in the cafe.
  3. I like the other chair better.
  4. I saw the man with glasses.
- We will explain in much more detail later in the course.

Course Organisation

Course structure and organization

- Lecturers: Adam Lopez, Edoardo Ponti, Shay Cohen, and lots of help!
  - In-person lectures three times a week. Recordings available shortly afterwards (but again: do not rely on this!)
  - Most lectures will have associated readings and a few quiz questions for you to check your understanding.
- Each week you will need to work through the readings and quizzes in addition to attending lectures. You will also (usually) have some further exercises.
  - Labs: practical computer-based work
  - Tutorial exercises: pencil-and-paper problems and questions.
  - These exercises are for your own learning. They will not count toward your final mark, but you are unlikely to get a good mark unless you engage with them!

Learning outcomes and assessment

Students who pass the course must demonstrate achievement of the 4 learning outcomes. That is, you will be able to...

1. identify, construct, and analyse examples of different kinds of ambiguity in natural language (e.g., ambiguity in part-of-speech, word sense, syntactic attachment). Explain how ambiguity presents a problem for computational analysis, and some of the ways it can be addressed.
2. describe and apply standard sequence and classification models; describe parsing and search algorithms for different levels of analysis (e.g. morphology, syntax, and semantics) and simulate each algorithm step-by-step with pen and paper.

Learning outcomes and assessment

3. for a range of NLP tasks, outline a processing pipeline for that task, including standard data sets, models, algorithms, and evaluation methods. Given a particular pipeline or part of the pipeline, identify potential strengths and weaknesses of the suggested dataset/method (including both technical and ethical issues, where appropriate), and provide examples to illustrate.
4. implement parts of the NLP pipeline with the help of appropriate support code and/or tools. Evaluate and interpret the results of implemented methods on natural language data sets.

Learning outcomes and assessment

To assess the learning outcomes, we will use:

- For LOs 1-3: a final exam (70%).
  - short factual answers, longer open-ended answers, problem-solving (maths, linguistics, alogithms).
- For LO 4: two assignments (15% each).
  - require some programming, but assessed on explanations and "lab-report" style write-ups.
  - Details in the Course Syllabus

How will you get help?

There is a lot to learn, but we will not leave you on your own. We will be doing our best to

- help you meet and learn together with other students, by putting you into pairs and groups for many of the course activities.
- give you opportunities to get help from course staff, by scheduling help hours and providing a class question forum. In the forum, you are encouraged to *answer* as well as ask questions. This helps you test your knowledge. We will monitor the forum and correct any misunderstandings.
- provide you with timely feedback on your learning in between major assessments (using automatic self-checks where possible, plus discussion in tutorial groups)

However, our learning platforms, schedule, and teaching support are mostly new this year, and there will undoubtedly be hiccups. So we ask for your patience and understanding with us too!

Course materials, tools, and communication

- We use Learn and OpenCourse as our primarly online portals. Check these regularly for course materials and announcements.
- We will use several tools to support students working together and the various activities I've mentioned (labs, tutorial groups, help hours, question forum, etc).
  - The Syllabus briefly describes each of these, and we will provide more detailed instructions as each activity gets going.
  - You will likely see several of them also in other courses.
  - A few require you to sign up for an account if you don't have one already. Please see the Week 1 checklist on Learn!

Background needed for this course?

- Know or currently learning Python.
- Background in Linguistics and prepared to learn maths (mainly probability) and algorithms
- Background in CS and prepared to learn linguistics (and maybe maths)

Advice/warnings

- This course is doable but very intensive if you:
  - have little programming/maths, or
  - have programming but little maths or weak English
- Start work early and try to find study/work partner(s) with other backgrounds or stronger English.
- Both ANLP and NLU+ focus on **language**, using **English examples**. If your English test scores are at the minimum required for Informatics degree, this course will be very hard.
- Other students: still fairly intense and scope for challenge. Don't underestimate the need to develop critical thinking and writing.

Quotes from course feedback forms

"What would you say to students interested in taking this course?"

- *Do everything that you are told to do/read, do not underestimate anything, devote a lot of time.*
- *It is a good course. Although it is very intensive, I did learn a lot more stuff than I expected. As long as you take advantage of all the learning resources provided and work hard on every assignment, you will definitely benefit a lot from it.*
- *Try hard on the assignments, but focus more on steadily preparing for the exam.*
- *It's a great course, but it's not a walk in the park, so be prepared to work hard.*
- *You'll learn a lot, but it is challenging.*

Relationship to other NLP courses

- ANLP is **required** if you want to take NLU+ in sem 2.
  - NLU+ builds on concepts in ANLP to cover advances, including lots about deep learning approaches.
  - ANLP course covers the linguistic, mathematical, and computational background needed first.
- Alternative text processing course: TTDS (20 pts, full year)
  - Focuses more on web search and shallow text processing
  - Less about the subtleties of language structure and meaning
  - More weight on practicals, including team project
  - Assumes more maths and programming background

What you should do next

Create structure and habits.

- Make a schedule. Plan out 12-13 hours/week to work on this class, at regular times.
- Read through the syllabus, which provides specific suggestions for deadlines during the week to complete each type of activity, and other tips for doing well in the course.
- Start working through the Week 1 checklist on Learn.