Natural Language Processing

Lecture 1: Introductions

Ross Deans Kristensen-McLachlan
rdkm@cas.au.dk

Plan for today

- Introducing ourselves
  - Who am I and who are you?
- Introducing the course
  - Course structure, *studieordning*, exam, etc
- Introducing NLP
  - What is it? And why won’t people shut up about it?

Introducing ourselves

- Tell me…
  - Your name (and any preferred name)
  - Any experiences you have of NLP (previous research, work experience)
  - Your main interest in cognitive science
  - And another interest you have!

Introducing the course

- Lectures & classrooms
- Exam format
- Academic goals
- Syllabus

Lectures & classrooms

Lectures & classrooms

- Lectures are generally more theoretical and conceptual
  - How do we understand this model?

Lectures & classrooms

- Lectures are generally more theoretical and conceptual
  - How do we understand this model?

- Classrooms are generally more practical and applied
  - How do we actually implement or use this model?

Lectures & classrooms

- Lectures are generally more theoretical and conceptual
  - How do we understand this model?
- Classrooms are generally more practical and applied
  - How do we actually implement or use this model?
- In both cases, the main source of content is GitHub
  - https://github.com/CHCAA-EDUX/NLP-E21
  - Brightspace will be kept synchronised

Exam format

- The exam is an individual oral exam based on a written synopsis. The duration is 30 minutes including the student's presentation of the synopsis project, followed by dialogue with the examinators and assessment.

- The synopsis can be done as an individual or group assignment. If it is done as a group assignment, the final product must (i) form a coherent text and (ii) be organized so that it is possible to make individual assessments of the students contributing. In other words, the contribution of each individual student to the whole assignment must be clearly delineated (excluding the introduction, conclusion and bibliography).

- A maximum of three students can take part in a group assignment.
  Length of individual synopsis: 4-7 standard pages (not including code and figures)
  Length of synopsis for 2 students: 8-14 standard (not including code and figures)
  Length of synopsis for 3 students: 12-21 standard (not including code and figures)

Academic goals

- Knowledge:
  - contrast different natural language processing methods in terms of their strengths and weaknesses in different use contexts
  - explain how formal analysis of natural language can provide insights into human cognition and behaviour
  - discuss ethical and philosophical issues connected to natural language processing software and technology.

Academic goals

- Skills:
    - identify relevant data sources for specific research and applied questions
    - correctly choose and apply tools for analysing natural language data.

Academic goals

- Competences:
  - critically reflect on and discuss theoretical and empirical implications of using natural language processing techniques
  - justify the choice between relevant methods and analyses used for specific research questions within the field of natural language processing
  - clear oral presentation of complex analyses.

Syllabus

Week	Session	Lecture	Classroom	Reading
36	1	Introductions	Work stack - Slack, UCloud, Github	--
	2	Simple text processing	Tokenization, word counts, collocation	*Hunston 2002, Chapters 1,2 4*
	3	From text processing to NLP	Lemmatization, part-of-speech, dependencies	*spaCy documentation, Enevoldsen et al 2021*
	4	Classification 1	Logistic regression and vectorization	*Jurafsky & Martin 2020, Chapter 5*
	5	Classification 2	Neural networks with pytorch	*Nielsen 2019, Chapter 1 (and throughout!)*
	6	Word embeddings	Named Entity Recognition	*Mikolov et al 2013; Pennington et al 2014*
BREAK	BREAK	BREAK	BREAK	BREAK
43	7	Language Modelling 1	Recurrent Neural Networks, Long short-term memory networks	*Urban & Gates 2020*
	8	Language Modelling 2	Attention	*Vaswani et al 2017; Lindsay et al 2020*
	9	BERT	Implementing (parts of) BERT	*Devlin et al 2019*
	10	***Project presentations***	Implementing (parts of) BERT	*Ettinger 2020; Rogers et al 2020*
	11	More BERT	Exploring layers	*Coenen et al 2019*
	12	Data bias and ethics	Augmentation and data bias	*van Miltenburg et al 2021*

Break

Any questions?

Introducing NLP

Introducing NLP

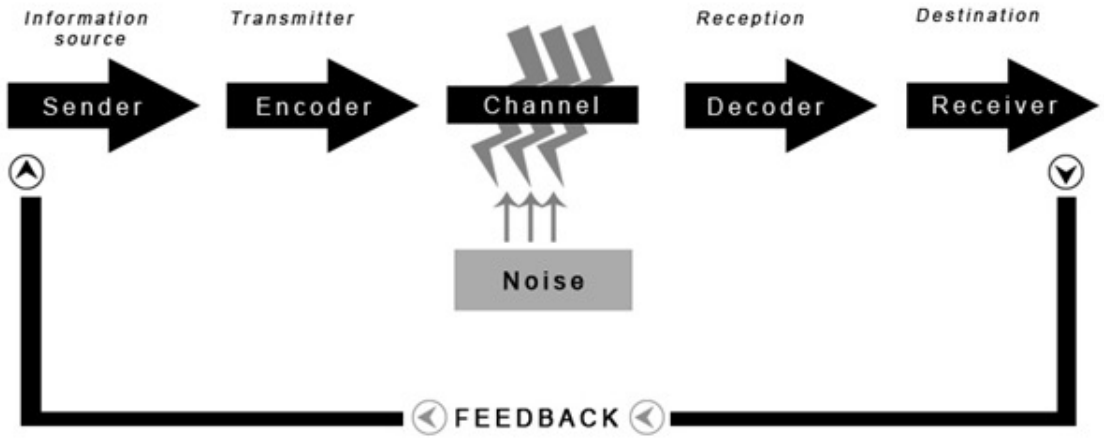
- The pre-history NLP
- NLP and cognitive science(s)
- What are we actually doing here?

The pre-history of NLP

- *Weaver's memorandum (1949)*
- *Chomsky's Syntactic Structures (1957)*
- *The first AI winter (Late 1970s)*
- *Statistical models (Late 80s-Early 90s)*

Weaver's Memorandum (1949)

- Weaver was a pioneer in information science
  - Along with Claude Shannon, he gives his name to the Shannon-Weaver model of communication
- In the late 40s, Weaver suggested that computers could be used for automated natural language translation
- In this short paper, Weaver made four proposals



SHANNON-WEAVER'S MODEL OF COMMUNICATION

Four proposals

- 1. Words can be disambiguated from *context*
- 2. Translation could be address as a problem of *formal logic*
- 3. Methods from *cryptography* could be useful for machine translation
- 4. *Linguistic universals* exist and these can be exploited to make translation simpler

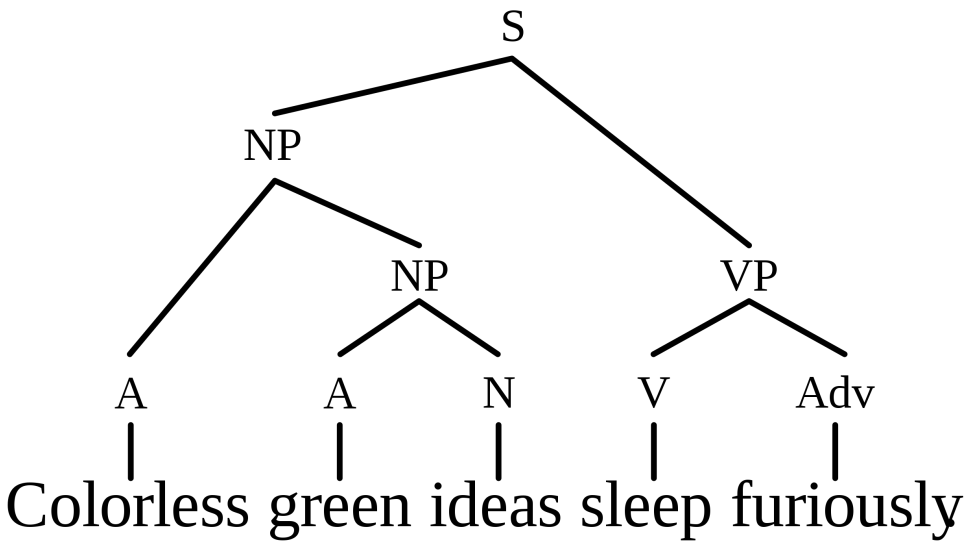
Some problems

- Perhaps unsurprisingly, this wasn't terribly successful
- It was cumbersome and time consuming to perform and the results were often nonsensical
- Lexical and grammatical ambiguities in natural language weren't taken into account
- One of the main problems with early experiments was that they lacked an adequate *theory of language*

Chomsky's *Syntactic Structures (1957)*

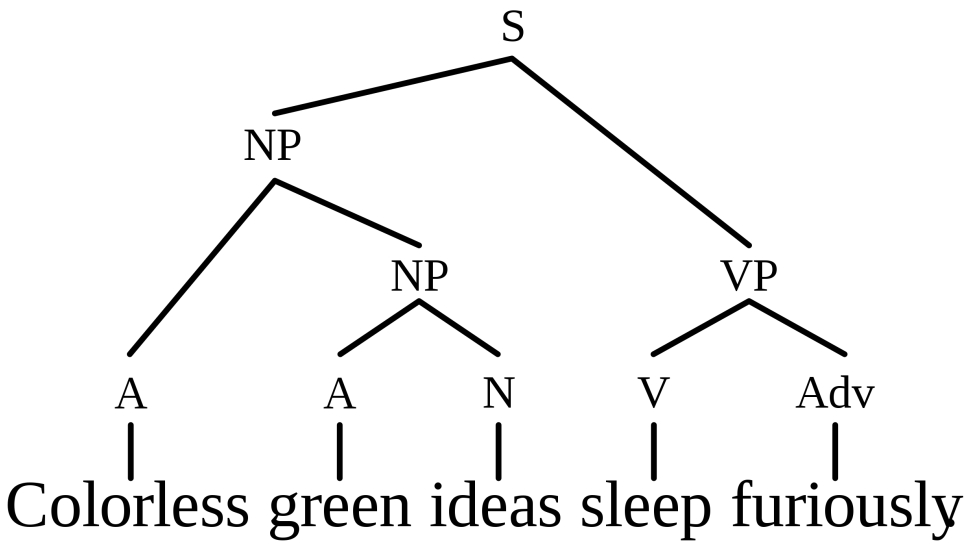
Chomsky's *Syntactic Structures (1957)*

- Chomsky argued that grammar was essentially a system of rules
- These rules generate exactly all possible combinations of *any* grammatical sentences in any given language
- The task of linguistics was to uncover and formalise this system of rules for any given language (and for Language generally)
- For this to be the case, Chomsky assumes…



Chomsky's *Syntactic Structures (1957)*

- 1. A grammatical sentence need not be included in any existing corpus
- 2. It need not be semantically meaningful
- 3. It does not need to be statistically likely



Chomsky's *Syntactic Structures (1957)*

"Despite the undeniable interest and importance of semantic and statistical studies of language, they appear to have no direct relevance to the problem of determining or characterizing the set of grammatical utterances. I think we are forced to conclude that grammar is autonomous and independent of meaning, and that probabilistic models give no particular insight into some of the basic problems of syntactic structure"

Chomsky (1957:17)

The First AI winter (1970s-ish)

- Chomsky insisted that grammar and semantics were separate. The goal of linguistics should be to model grammar
- Naturally, this resulted in a huge effort to computationally represent meaning. Many argued that Chomsky was *too* syntactic
- Around the same time, Minksy & Papert (1969) wrote what was - at the time - considered a crushing blow to the connectionist paradigm in AI (i.e. neural nets)
- It's difficult to summarise everything that went on in this period but a good example is Schank & Abelson's (1977) Script Theory

Schank & Abelson (1977)

- Schank & Abelson (and others) began to develop a cognitive psychological theory based around the notion of scripts

- Scripts are structured, organised representation of a stereotyped sequence of events

- These structure scripts allow AI to understand aspects of language which go beyond syntactic processing and approach the level of *discourse*

- Scenario mapping theory, Sanford & Garrod (1981);
- Mental models, Johnson-Laird (1983);
- Rhetorical processing framework, Sanford & Emmott (2012)

Script: Restaurant
Entry Conditions: Customer is hungry.
Customer has money.
Scenes:

1. Entering

Customer goes into restaurant. (E_1)
Customer looks around. (E_2)
Customer decides where to sit. (E_3)
Customer goes to a table and sits down. (E_4)

2. Ordering

Customer picks up a menu. (E_5)
Customer decides on food. (E_6)
Customer orders food from waiter. (E_7)
Waiter tells cook the order. (E_8)
Cook prepares food. (E_9)

3. Eating

Cook gives food to waiter. (E_{10})
Waiter gives food to customer. (E_{11})
Customer eats food. (E_{12})

4. Exiting

Waiter writes out check. (E_{13})
Waiter brings check to customer. (E_{14})
Customer gives tip to waiter. (E_{15})
Customer goes to cash register. (E_{16})
Customer gives money to cashier. (E_{17})
Customer leaves restaurant. (E_{18})

Statistical models (Late 80s, early 90s)

- Scripts/schemata/scenarios/frames are (usually) context specific

Statistical models (Late 80s, early 90s)

- Scripts/schemata/scenarios/frames are (usually) context specific

- From an AI or NLP perspective, this means we need to have some kind of representation of every possible context we are likely to encounter

Statistical models (Late 80s, early 90s)

- Scripts/schemata/scenarios/frames are (usually) context specific
- From an AI or NLP perspective, this means we need to have some kind of representation of every possible context we are likely to encounter
- From an engineering perspective, computers should *learn* from language data, rather than us hand-coding knowledge representations ahead of time

Statistical models (Late 80s, early 90s)

"Rather than starting off by dividing sentences into grammatical and ungrammatical ones, we instead ask, *What are the common patterns that occur in language use? […] While practical utility is something different from the validity of a theory, the usefulness of statistical models of language tends to confirm that there is something right about the basic approach"*

Manning & Schütze (1999:4)

NLP and cognitive science(s)

NLP and cognitive science(s)

- It’s… complicated

NLP and cognitive science(s)

- It’s… complicated

- NLP does not strictly commit to any cognitive theory
  - Broadly speaking ‘connectionist’. But what does that mean?

NLP and cognitive science(s)

- It’s… complicated
- NLP does not strictly commit to any cognitive theory
  - Broadly speaking ‘connectionist’. But what does that mean?
- Neither NLP or cognitive science are monolithic practices

NLP and cognitive science(s)

- It’s… complicated
- NLP does not strictly commit to any cognitive theory
  - Broadly speaking ‘connectionist’. But what does that mean?
- Neither NLP or cognitive science are monolithic practices
- NLP and computational linguistics share a similar domain but different goals
  - Think of the physicist vs the engineer

So what are we doing here?

So what are we doing here?

- Well, we’re doing natural language processing!

So what are we doing here?

- Well, we're doing natural language processing!
- We'll primarily be working in the dominant paradigm of deep learning and neural language models

So what are we doing here?

- Well, we're doing natural language processing!
- We'll primarily be working in the dominant paradigm of deep learning and neural language models
- We'll be learning how to build these models from scratch *and* how can we fine tune these models for specific purposes

Questions?

Bibliography

Chomsky, N. (1957). *Syntactic Structures.* The Hague: Mouton de Gruyter.

Johnson-Laird, P.N. (1983). *Mental Models*. Cambridge: Cambridge University Press.

Manning, C.D. & Schütze, H. (1999). *Foundations of Statistical Natural language Processing*. Cambridge, MA: MIT Press.

Minksy, M. & Papert, S. (1969). *Perceptrons.* Cambridge, MA: MIT Press.

Sanford, A.J. & Emmott, C. (2012). *Mind, Brain and Narrative*. Cambridge: Cambridge University Press.

Sanford, A. & Garrod, S.C. (1981). *Understanding Written Language: Explorations in Comprehension Beyond the Sentence*. Chichester: John Wiley & Sons.

Schank, R.C. & Abelson, R.P. (1977). *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures.* Hillsdale, NJ: Psychology Press.

Weaver, W. (1949). 'Translation', Reproduced in: *Locke, W.N.; Booth, D.A., eds. (1955), Machine Translation of Languages.* Cambridge, MA: MIT Press, 15–23.