IN4080 – 2020 FALL

NATURAL LANGUAGE PROCESSING

Jan Tore Lønning

Today

- Part 1: Course overview
  - What is this course about?
  - How will it be organized?
  - Interactive zoom

- Part 2: ”Looking at data”:
  - Descriptive statistics
  - Some language data
  - Video lectures

Name game

- Computational Linguistics
  - Traditional name, stresses interdisciplinarity
- Natural Language Processing
  - Computer science/AI/NLP
  - ”Natural language” a CS term
- Language Technology
  - Newer term, emphasize applicability
  - LT today is not SciFi (AI), but part of everyday app(lication)s
- The terms have different historical roots
  - Today: NLP=Computational Linguistics, restricted to written language
  - LT = NLP + speech (No speech in this course)

Megatrends

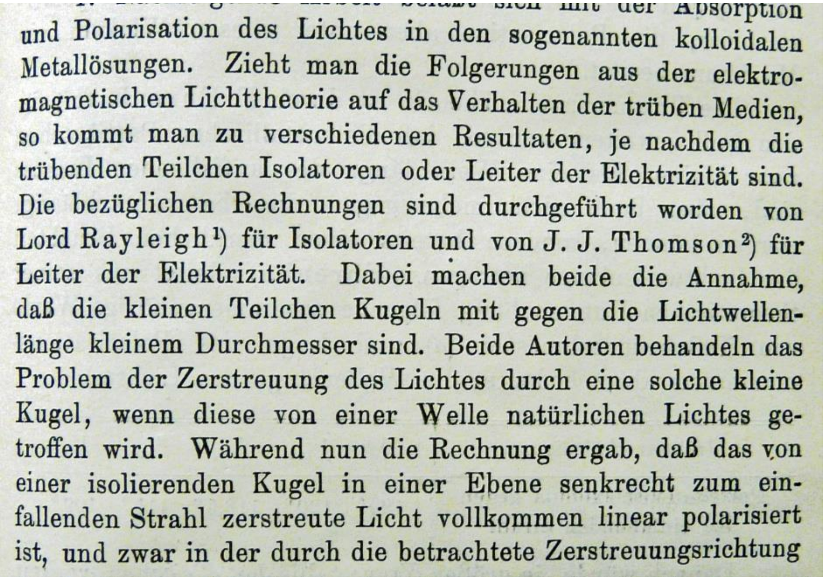
Natural Language Processing

Artificial Intelligence AI
- Machine learning
  - Deep learning

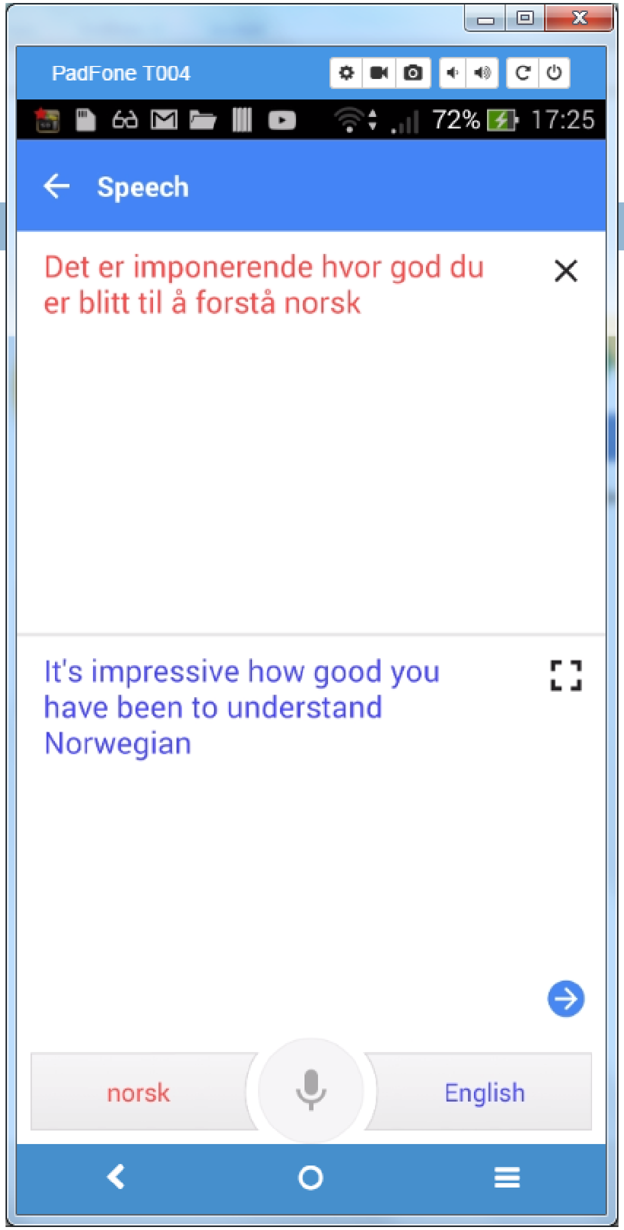
"Data science"
Big data
(WWW)

Language technology: examples

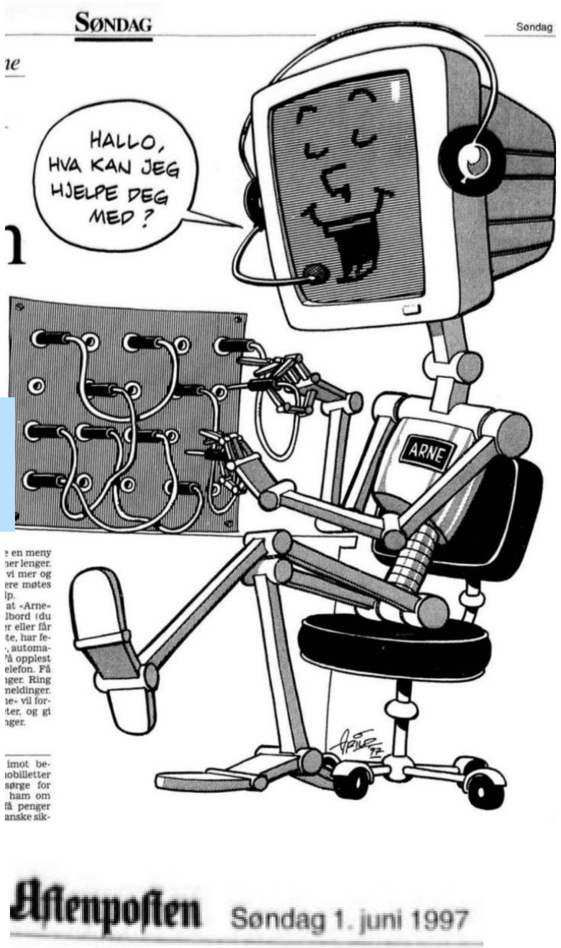
1. Speech ⟷ text


und Polarisation des Lichtes in den sogenannten kolloidalen Metallösungen. Zieht man die Folgerungen aus der elektromagnetischen Lichttheorie auf das Verhalten der trüben Medien, so kommt man zu verschiedenen Resultaten, je nachdem die trübenden Teilchen Isolatoren oder Leiter der Elektrizität sind. Die bezüglichen Rechnungen sind durchgeführt worden von Lord Rayleigh[1]) für Isolatoren und von J. J. Thomson[2]) für Leiter der Elektrizität. Dabei machen beide die Annahme, daß die kleinen Teilchen Kugeln mit gegen die Lichtwellenlänge kleinem Durchmesser sind. Beide Autoren behandeln das Problem der Zerstreuung des Lichtes durch eine solche kleine Kugel, wenn diese von einer Welle natürlichen Lichtes getroffen wird. Während nun die Rechnung ergab, daß das von einer isolierenden Kugel in einer Ebene senkrecht zum einfallenden Strahl zerstreute Licht vollkommen linear polarisiert ist, und zwar in der durch die betrachtete Zerstreuungsrichtung

2. Machine translation



3. Dialogue systems



4. Sentiment analysis and opinion mining

Sentiment/opinion mining:

- Do consumers appreciate more sugar in the soda?
- Do (my core voters) like my last Twitter outburst?
- How will the stock prices develop?
- Is there a danger of a revolt in country X?

- Personalization:
  - Adds
  - News

5. Text analytics

- Goal, example IBM's Watson system:
- Read medical papers + records:
  - Propose diagnoses
  - Propose treatments

+

- Similarly in other domains:
  - Oil & Gas
  - Legal domain

6. NLP applications – more examples

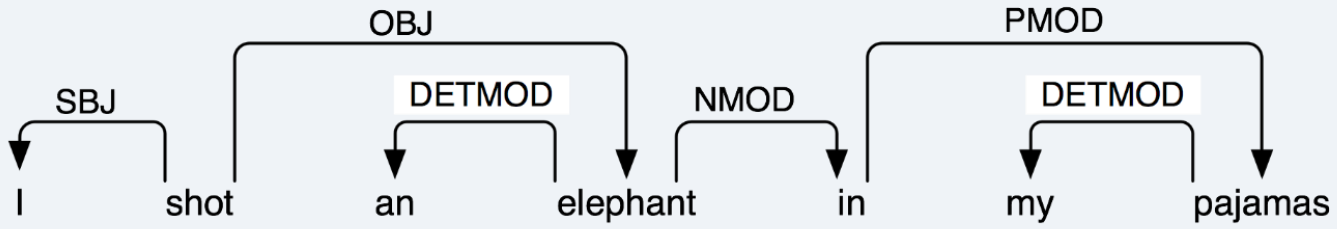
- Intelligence
- Surveillance:
  - How does NSA manage to read all those e-mails?
- User content moderation
- Election influence

What?

What

- https://www.uio.no/studier/emner/matnat/ifi/IN4080/index.html
- Follow steps in bottom-up data-driven text systems
- Learn to set-up and carry out experiments in NLP:
  - Machine learning
  - Evaluation
  - in-depth knowledge of at least one application
- Dialogue system (October)
  - "…in-depth knowledge of at least one [NLP] application…"
- In addition
  - Ethics in NLP

Some steps when processing text

Split into sentences	Obama says he didn't fear for 'democracy' when running against McCain, Romney.
Tokenize (normalize)	| Obama | says | he | did| not | fear | for | ' | democracy | ' | when | running | against | McCain | , | Romney | .
Tag	Obama_N says_V he_PN did_V not_ADV fear_V …
Lemmatize	Says_V → say_V, did_V → do_V, running_V → run_V …
Parsing (dependency)	
Coreference resolution	Obama says he did not …..
Semantic relation detect.	Fear(Obama, Democracy) Run_against(Obama, McCain),..
Negation detection	… did not fear … → Not(Fear(Obama, Democracy))

The two cultures (up to the 1980s)

Symbolic

- 1956 →
- Sub-cultures
  1. AI (NLU)
     - McCarthy, Minsky → SHRDLU ('72)
  2. Formal Linguistics/Logic
     - Chomsky
       - automata, formal grammars
     - + Logic in the 80s
     - LFG, HPSG
  3. Discourse, pragmatics

Stochastic

- Information theory, 1940s
- Statistics
- Electrical engineering
- Signal processing

Trends the last 30 years

- 1990s: combining the cultures
  - methods from speech adopted by NLP
    - division of labor between methods
    - stochastic components in symbolic models, e.g. statistical parsing
  - (larger) text corpora
  - Jurafsky and Martin, SLP, 2000
- 2000s:
  - More and more machine learning in NLP, at all levels
  - Examples and corpora
  - Rethinking the curriculum and the order in which it is taught
  - J&M, 2. ed, 2008

Example:
machine translation systems that are trained on earlier translated texts

Currently

- 2010s Deep learning
  - ML with multi-layered Neural Networks
  - Revolution, in particular for
    - Image recognition
    - Speech
  - Entered into all parts of NLP
    - Key: "Word embeddings"

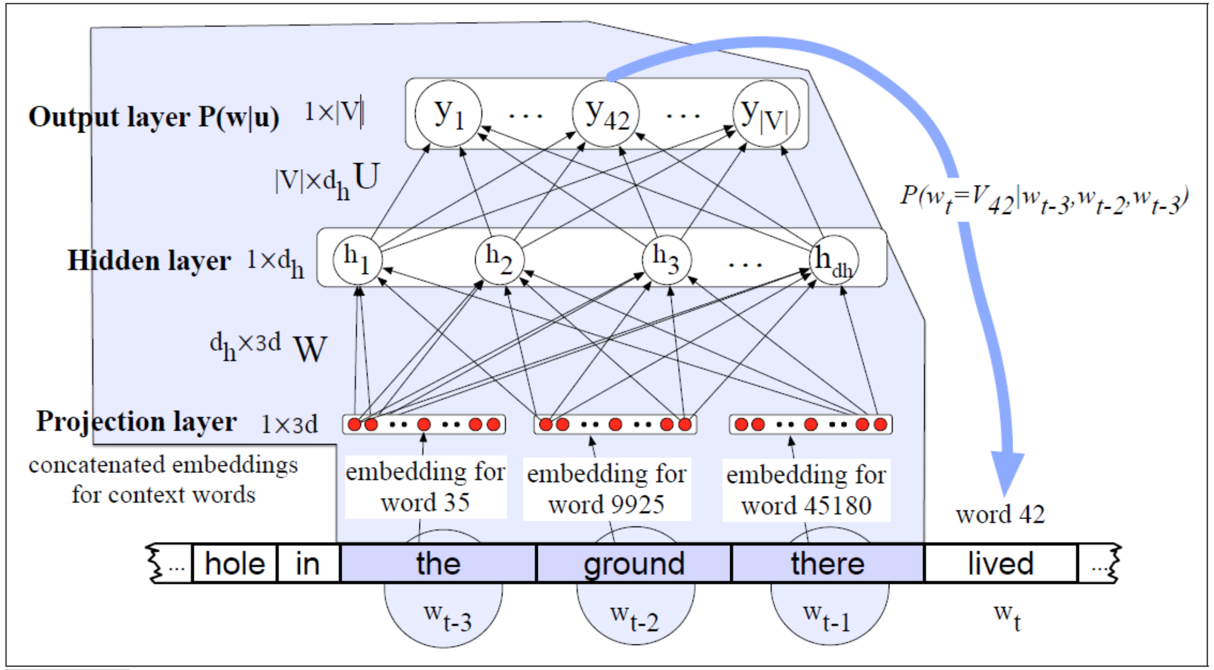


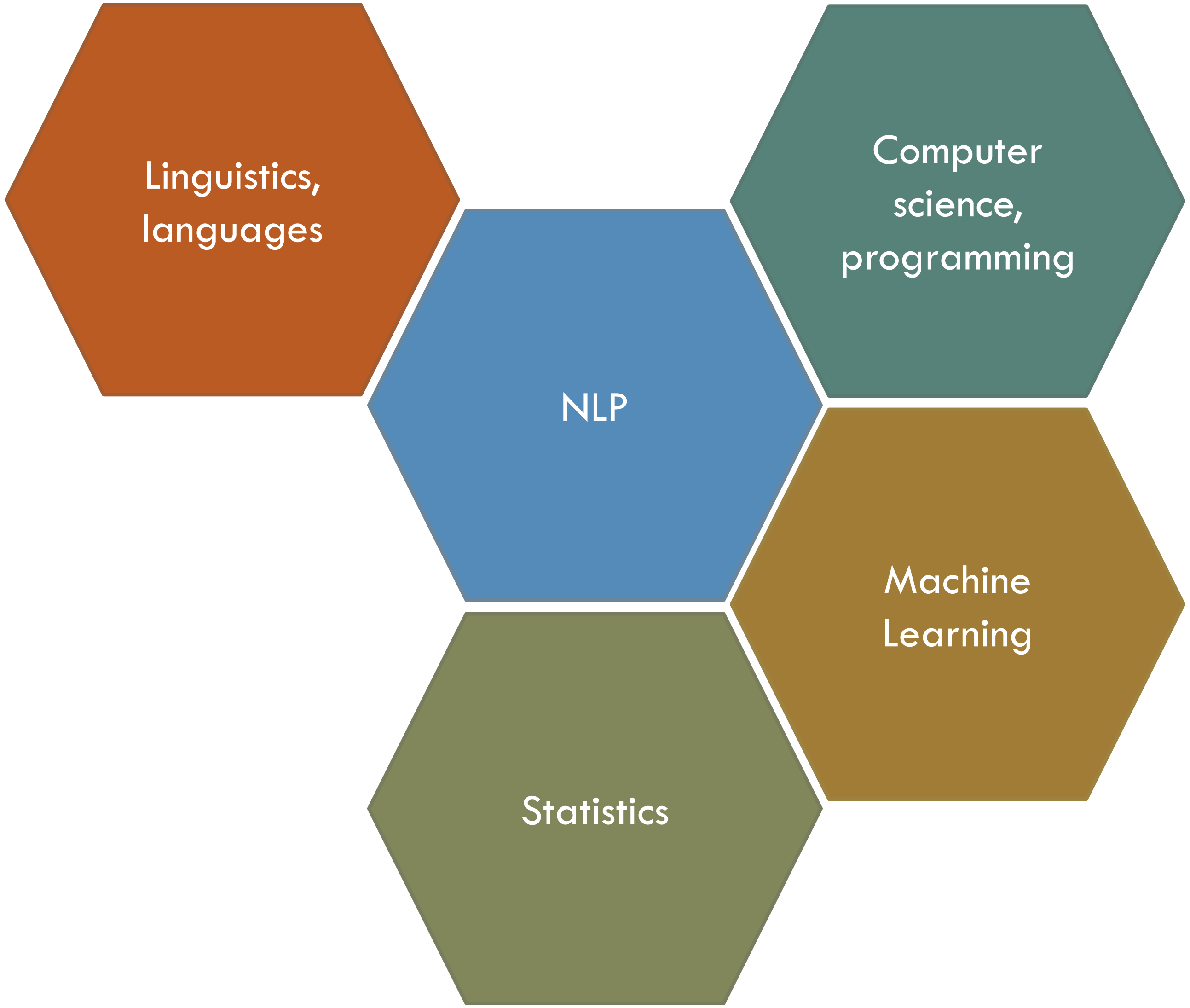
Figure 9.1 A simplified view of a feedforward neural language model moving through a text. At each time step t the network takes the 3 context words, converts each to a d-dimensional embedding, and concatenates the 3 embeddings together to get the $1 \times Nd$ unit input layer x for the network.

DL and IN4080

- Should we jump directly to deep learning?
- We will (initially) focus on simpler models.
- Most tasks are independent of learning algorithm, and can be easier understood using simpler models
- For several tasks, traditional ML is still compatible

- The inner workings of Deep learning in NLP is the topic in "IN5550 Neural Methods in NLP“, spring 2021

NLP is based on



Why statistics and probability in NLP?

1. "Choose the best"

(=the most probable given the available information)

- *bank* (Eng.) can translate to b.o. *bank* or *bredd* in No.
  - Which should we choose?
  - What if we know the context is "*river bank*"?
- *bank* can be Verb or Noun,
  - which tag should we choose?
  - What if the context is *they bank the money* ?
- A sentence may be ambiguous:
  - What is the most probable parse of the sentence?

Use of probabilities and statistics, ctd.:

2. In constructing models from examples (ML):
   - What is the best model given these examples?

3. Evaluation:
   - Model1 is performing slightly better than model 2 (78.4 vs. 73.2), can we conclude that model 1 is better?
   - How large test corpus do we need?

How?

Syllabus (online)

- Lectures, presentations put on the web
- Jurafsky and Martin, *Speech and Language Processing, 3.ed.*
  - In progress, edition of Oct. 2019
- Articles from the web
- In addition
  - Some selections from
    - S. Bird, E. Klein and E. Loper: *Natural Language Processing with Python*
    - available on the web, python 3 ed.
  - Probabilities and statistics (some book or)
    - www.openintro.org/stat/textbook.php

Challenges for a master's course like this

- You have different backgrounds:
  - Some are familiar with some NLP from e.g. IN2110
  - Some are familiar with simple probabilities and statistics, some are not
  - Some are familiar with Machine Learning
  - Some are familiar with Language and linguistics
- For teaching:
  - You might have heard some of it before
  - You might experience a step learning curve on other parts
- For you:
  - Concentrate on the parts with which you are less familiar

Schedule

- Lectures: Mondays 10.15-12
  - Room Java (34 seats)
  - Screencasts distributed after lecture
- Lab sessions: Tuesdays 10.15-12
  - Room: Fortress 3468, (18 seats)
  - No screencast
  - Booking system
- Some sort of zoom-group
- 3 mandatory assignments (oblig.s)
  - Weeks 37, 40, 43
- Written exam
  - Wednesday 2 December

PadLet for QAs
No Piazza or Slack (GDPR)

Tomorrow

- Tutorial on probabilities
- 10.15 Fortress
- Sign up
- Regular groups start 25.8

Background knowledge

- Please fill in:
- https://nettskjema.no/a/157223