IN1140: Introduksjon til språkteknologi

Forelesning #1

Erik Velldal & Lilja Øvrelid

Universitetet i Oslo

20 august 2025



Tema for i dag

- Introduksjon
- Hva er språkteknologi?
- Hva er IN1140?
- Praktiske detaljer
  - Undervisningsform
  - Grupper
  - Obliger
  - Lærebøker
  - Kontakt
  - m.m.



Screencasts + fysiske gruppetimer

UNIVERSITETET I OSLO

NO EN Meny

← Studier ← Emner ← Matematikk og naturvitenskap ← Informatikk ← IN1140 ← Høst 2025

Beskjeder

Obliger

Ukesplan

IN1140 - Høst 2025

Beskjeder

Publisert 19. aug. 2025 13:11

Obliger høsten 2025

Emnesiden er nå oppdatert med viktig informasjon om obligatoriske innleveringer i IN1140, viktig at alle setter seg inn i dette. Merk at obligene skal leveres som **gruppearbeid**, med 2–3 studenter per gruppe.

Publisert 13. aug. 2025 09:51

Undervisningsform høsten 2025

IN1140 går for siste gang denne høsten og gjennomføres som **selvstudium**. Det vil bli gitt fysisk forelesning i emnet kun to ganger (ved semesterstart og slutt) og ellers vil undervisningen basere seg på opptak fra tidligere år.

Hei

Emneansvarlige

- Lilja Øvrelid
- Erik Velldal
- Fra språkteknologigruppen (LTG), ved seksjon for maskinlæring (MLS)
- I forelesningsopptakene vil dere høre fra flere LTG-medlemmer, som Sondre Wold or Egil Rønningstad.

Gruppelærer

- Håkon Liltved Hyrve

Hva er språkteknologi?

- Mål: å få datamaskiner til å 'forstå' naturlige språk.
- Aka:
  - computational linguistics (datalingvistikk)
  - language technology
  - language engineering
  - natural language processing (NLP)



Språkteknologi og tverrfaglighet

NLP er et tverrfaglig felt

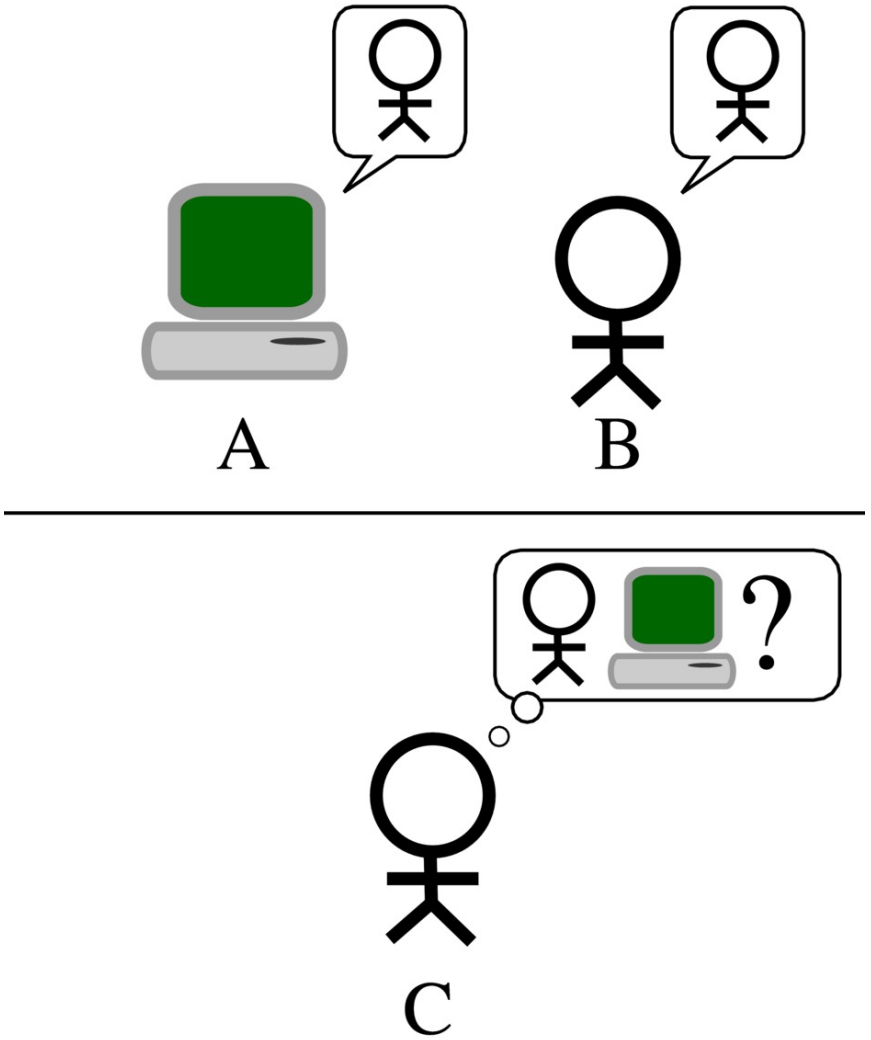
- Lingvistikk
- Informatikk
- Statistikk
- Maskinlæring
- Logikk, Filosofi, Psykologi, ...



- Del av det bredere feltet kunstig intelligens (AI).

Turingtesten

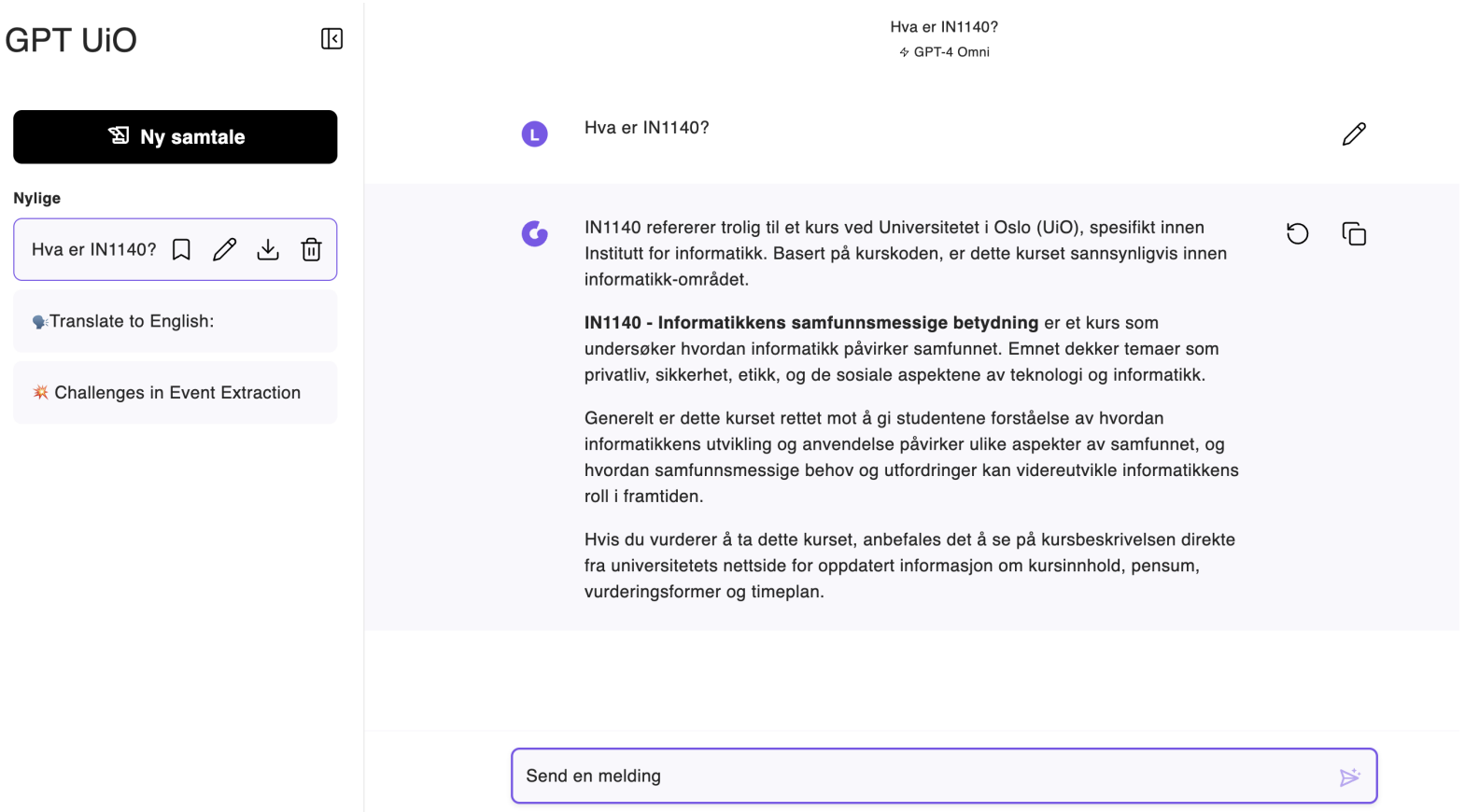
- Alan Turing i 1950:
- *I propose to consider the question, 'Can machines think?'*
- Definisjonsspørsmål. Skulle avgjøres ved Turingtesten.



Eksempler på språkteknologi?



Tekstgenerering



(https://www.uio.no/tjenester/it/ki/gpt-uio/)

Språkforståelse er vanskelig!

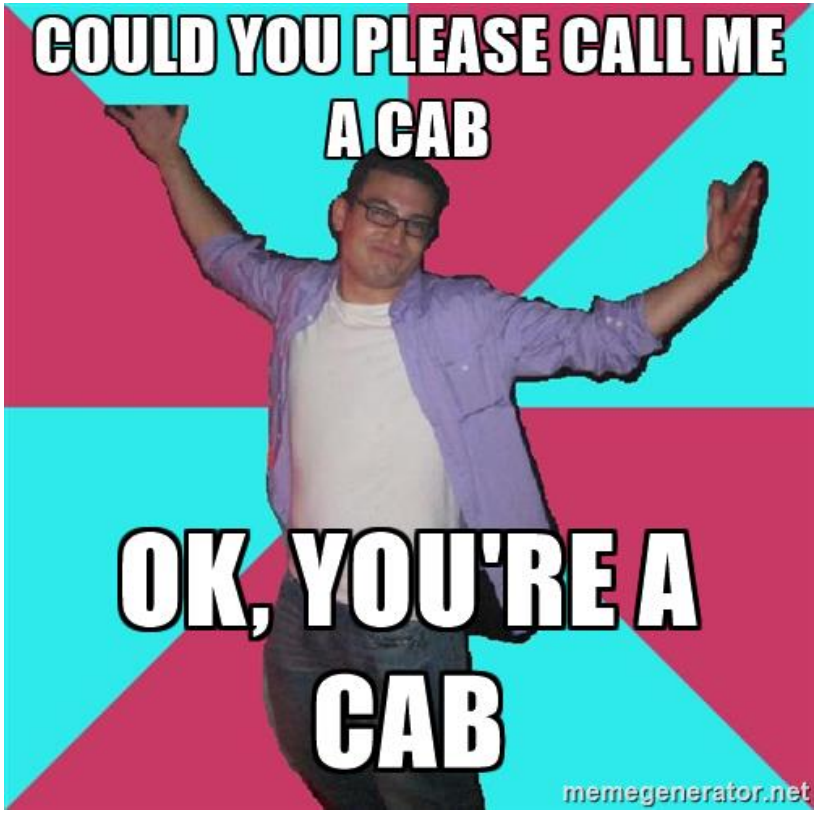
The main lesson of thirty-five years of AI research is that the hard problems are easy and the easy problems are hard.

Steven Pinker, *The language instinct*, 1994

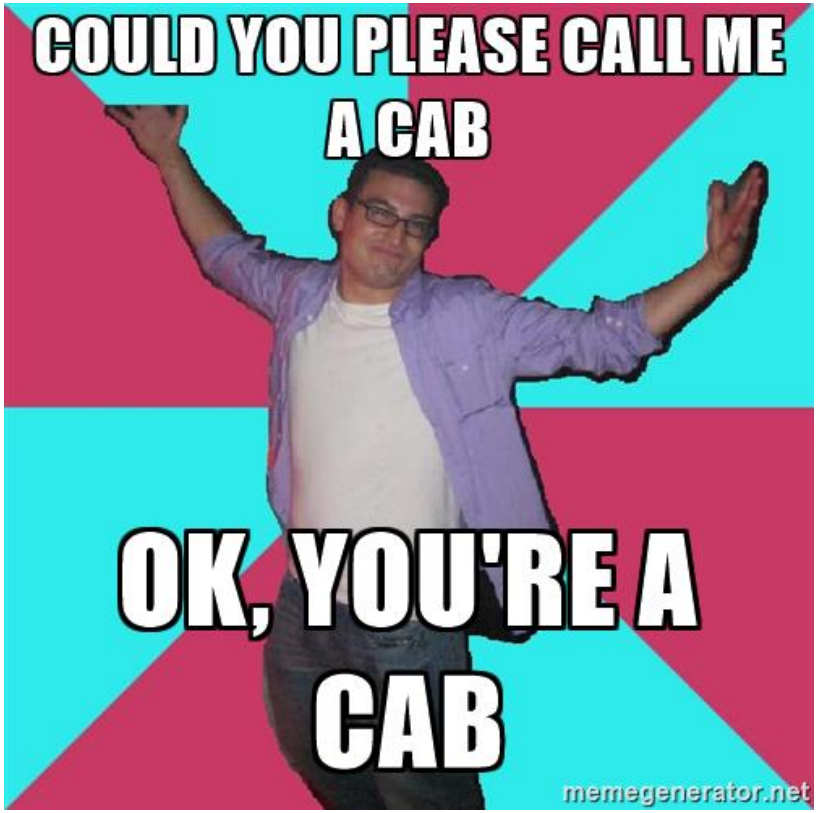
Et eksempel på "Moravec's paradoks" (Hans Moravec, 1988)

Hvorfor er språkforståelse utfordrende?

Hvorfor er språkforståelse utfordrende?

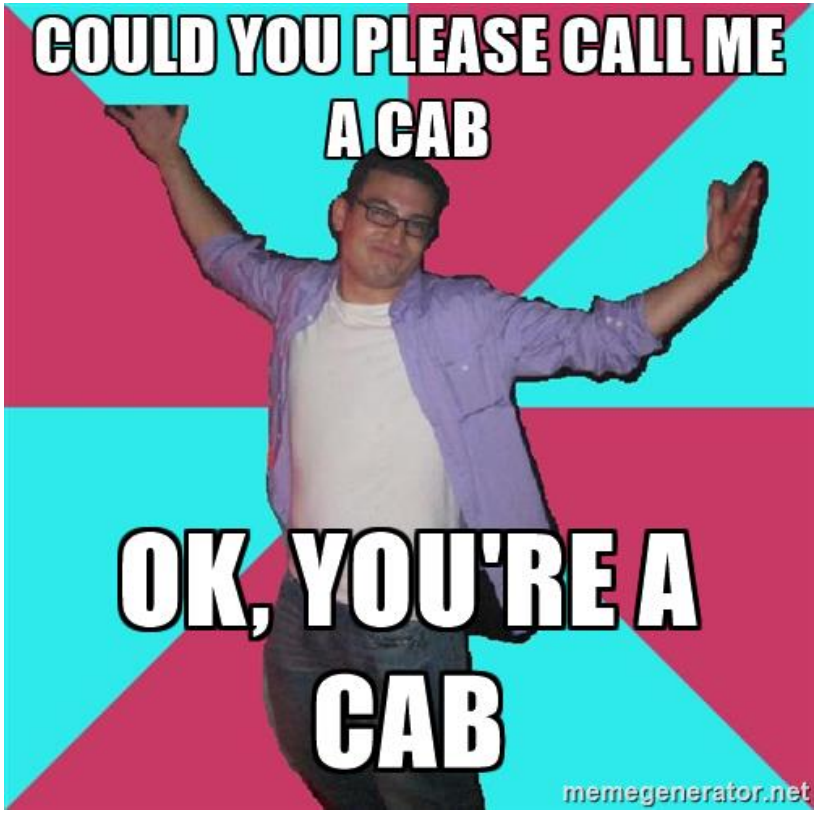


Hvorfor er språkforståelse utfordrende?



- Språk er vagt, ulike tolkninger mulig.
- Flertydighet overalt.
- Gir kompakt kommunikasjon:
- Samme uttrykk kan brukes i ulike kontekster.

Hvorfor er språkforståelse utfordrende?



- Språk er vagt, ulike tolkninger mulig.
- Flertydighet overalt.
- Gir kompakt kommunikasjon:
- Samme uttrykk kan brukes i ulike kontekster.

- Flertydighetene er stort sett usynlige for oss, vi finner den intenderte tolkningen nærmest ubevisst.
- For maskiner er det motsatt: lett å finne alle mulige tolkninger, men vanskelig å se hvilken som er riktig.

Eksempel: Flertydighet på ordnivå

- Norsk: *rett*.
- Engelsk: *?*

Eksempel: Flertydighet på ordnivå

- Norsk: *rett*.
- Engelsk: *?*
- Flertydig ift betydning + ordklasse (verb, subst., adj., adv.).
- Vi trenger kontekst for å avgjøre.

Eksempel: Flertydighet på ordnivå

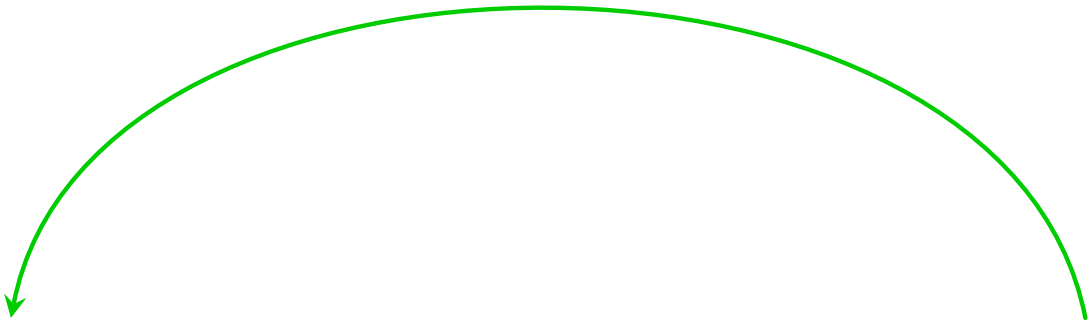
- Norsk: *rett*.
- Engelsk: *?*
- Flertydig ift betydning + ordklasse (verb, subst., adj., adv.).
- Vi trenger kontekst for å avgjøre.

avgrenset av en rett linje tvers over kanalen	*straight*
Hva er rett svar?	*correct, right*
lovbestemt rett til innsyn	*right*
Denne rett avsa enstemmig dom i saken 4. juli 1980	*court*
Norsk rett tilpasses EUs regelverk	*law*
Vennligst rett disse prøvene!	*grade, correct*
Det bar rett i fengsel	*directly, straight*
De spiste en deilig rett av grønnsaker.	*meal, dish*
han var rett utenfor, rett nå	*just*
Slikt skjer rett som det er.	må omskrives

Flertydighet i referanse

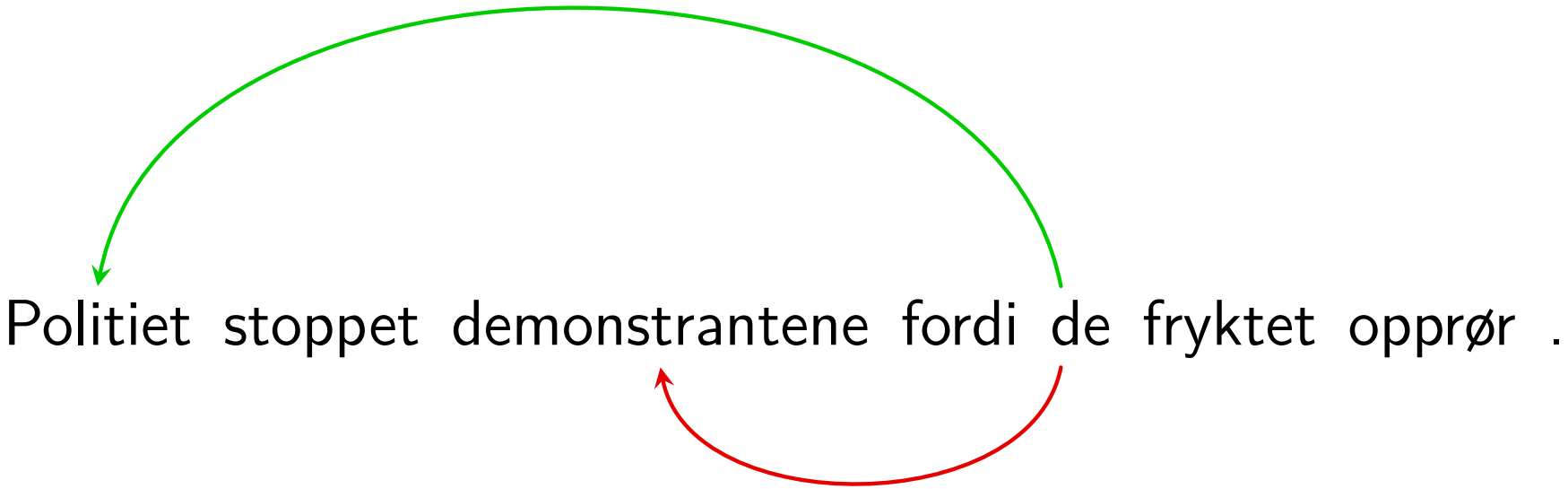
Politiet stoppet demonstrantene fordi de fryktet opprør .

Flertydighet i referanse

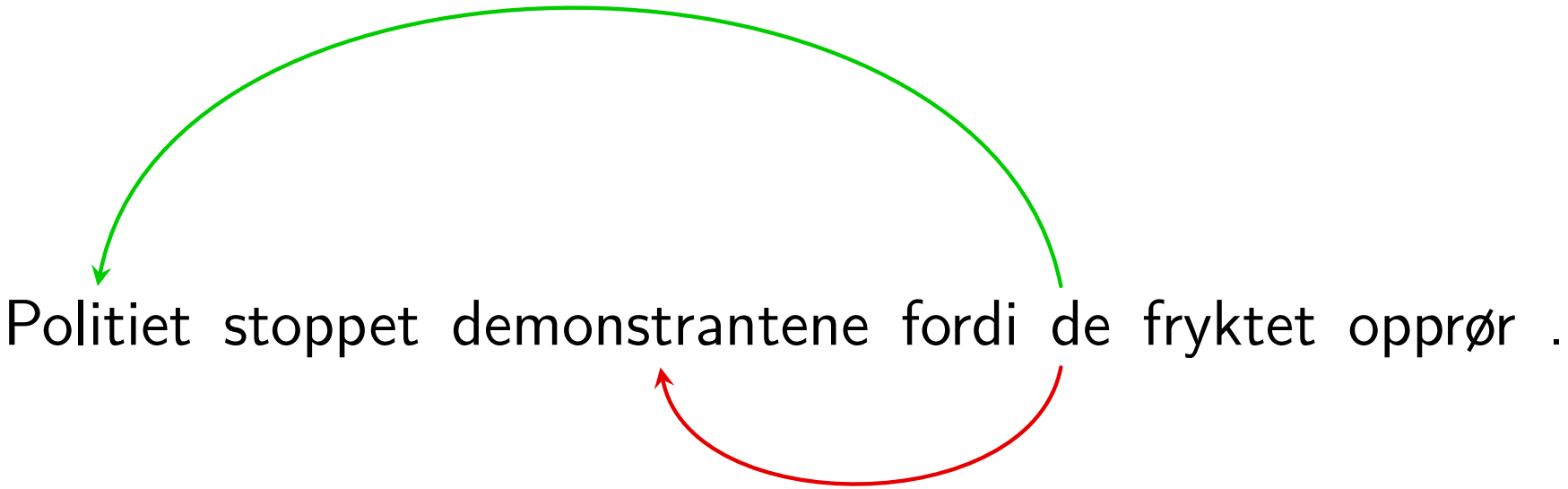


Politiet stoppet demonstrantene fordi de fryktet opprør .

Flertydighet i referanse

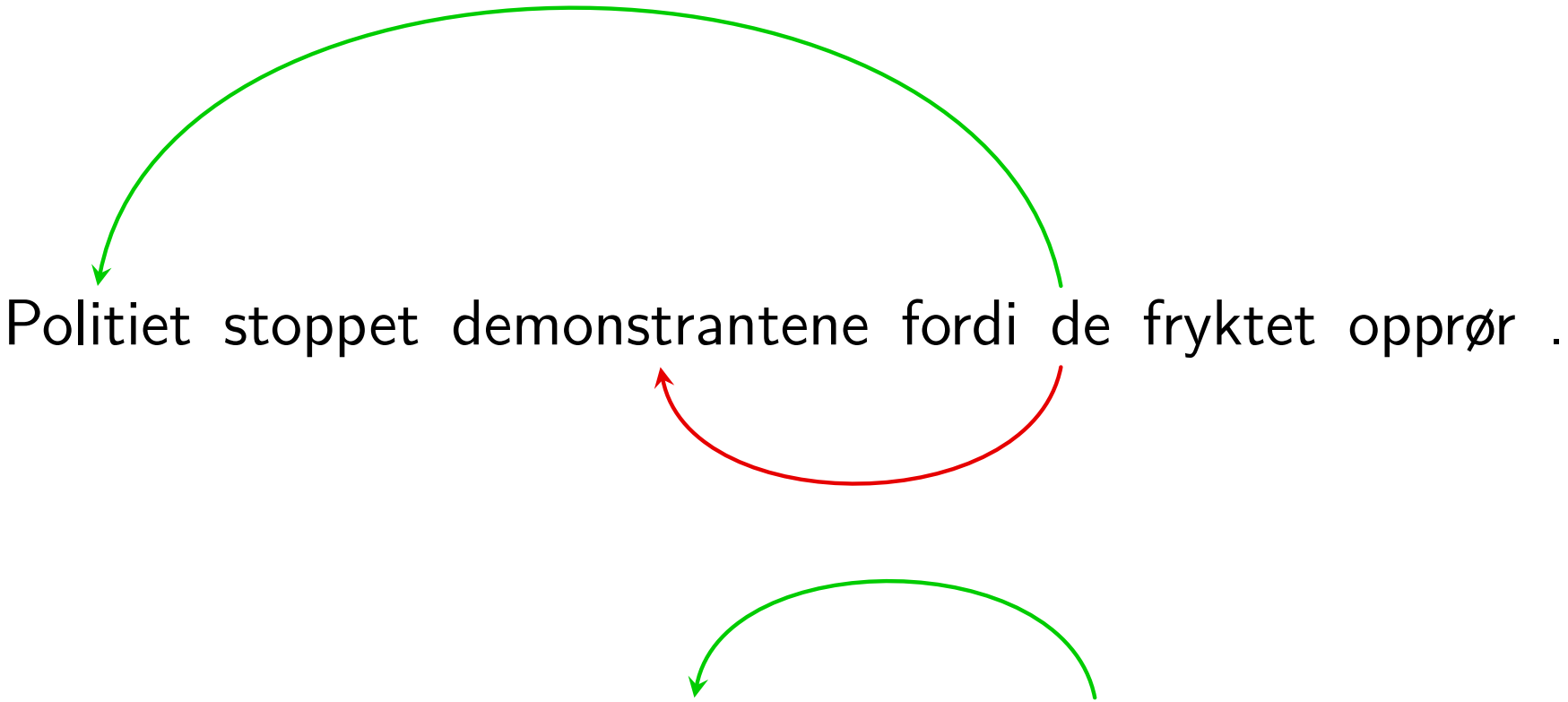


Flertydighet i referanse

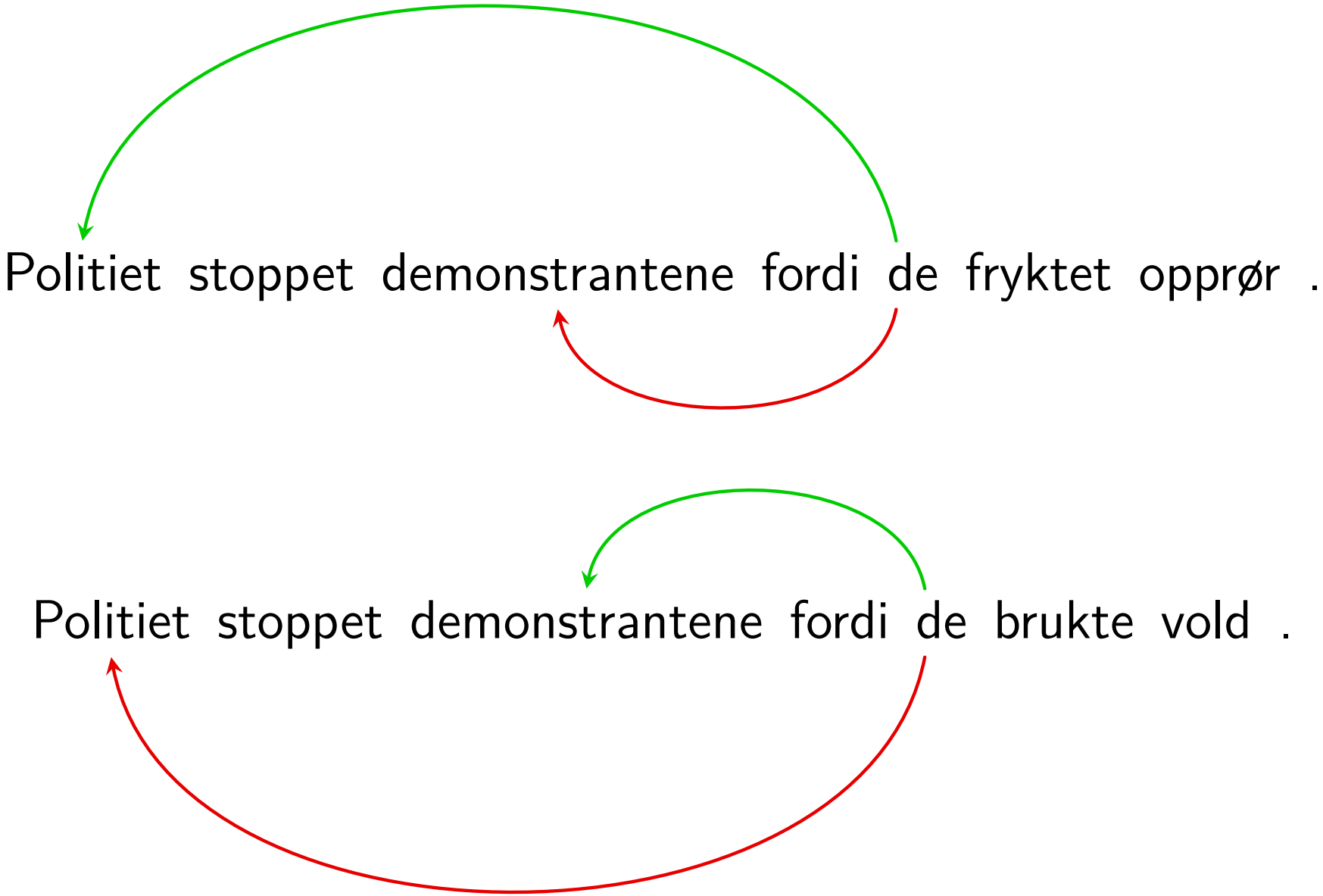


Politiet stoppet demonstrantene fordi de brukte vold .

Flertydighet i referanse



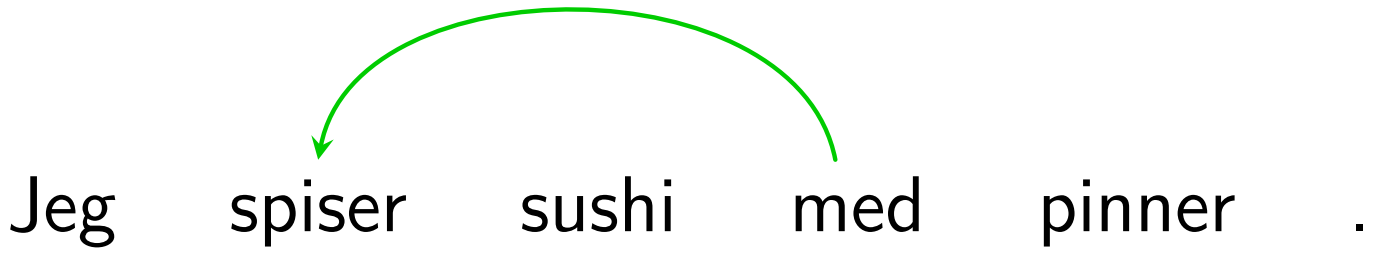
Flertydighet i referanse



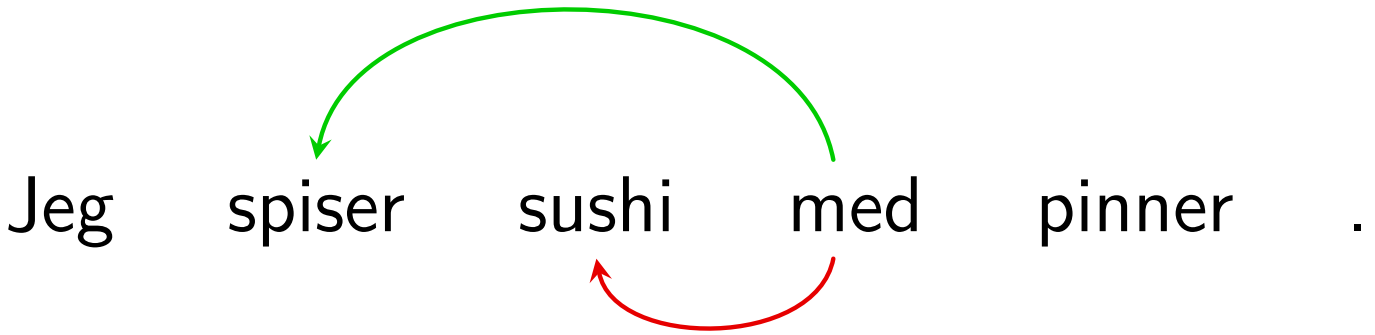
Flertydighet på setningsnivå

Jeg spiser sushi med pinner .

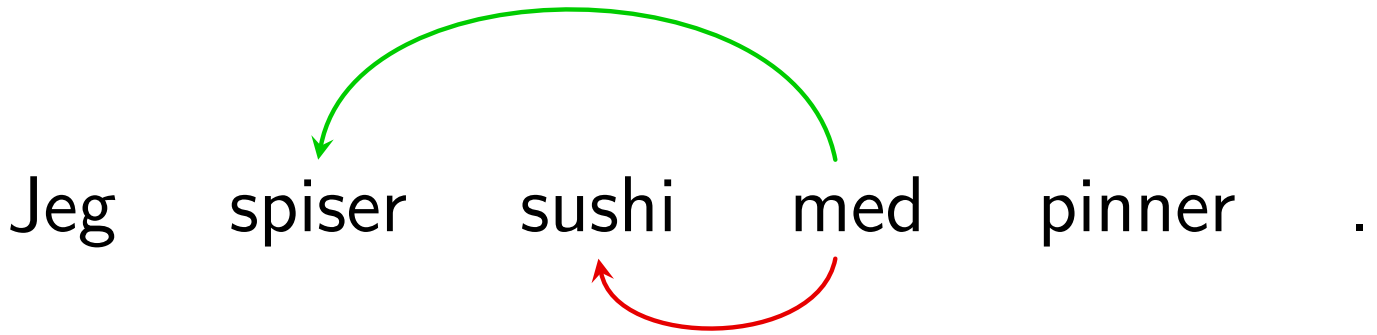
Flertydighet på setningsnivå



Flertydighet på setningsnivå

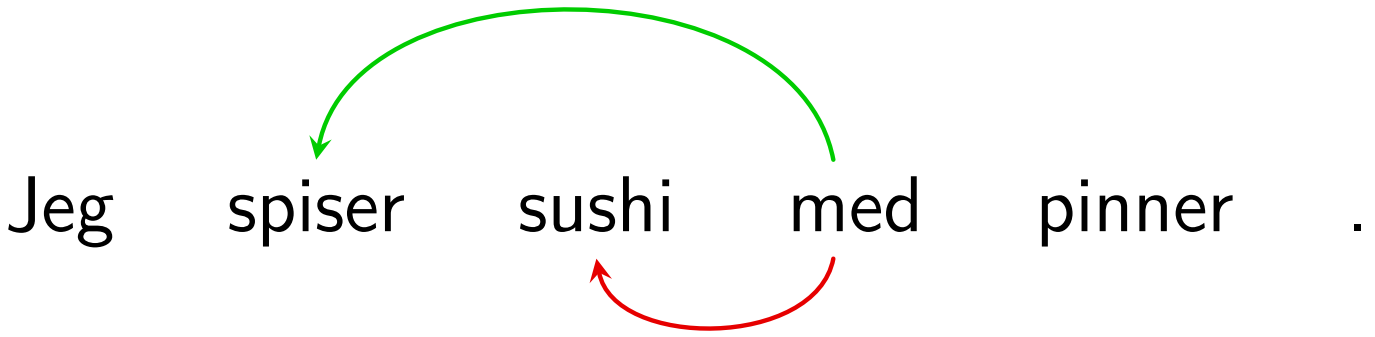


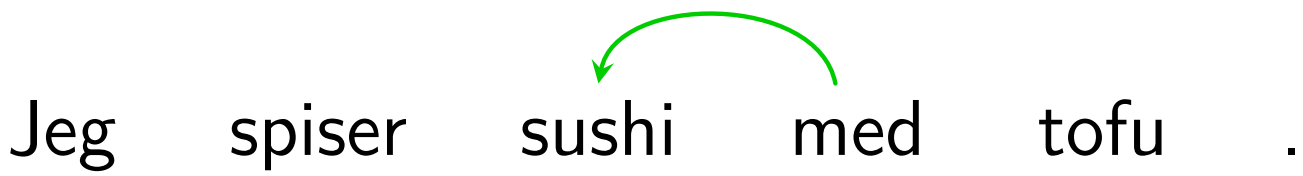
Flertydighet på setningsnivå



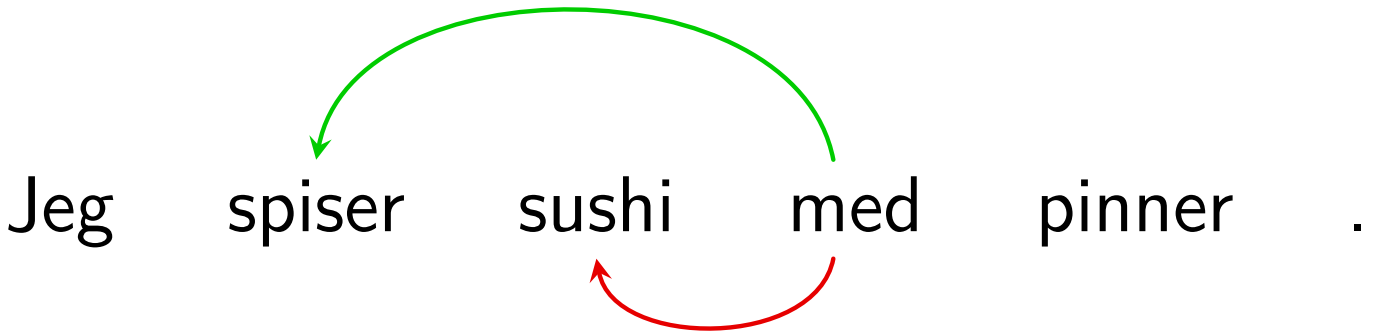
Jeg spiser sushi med tofu .

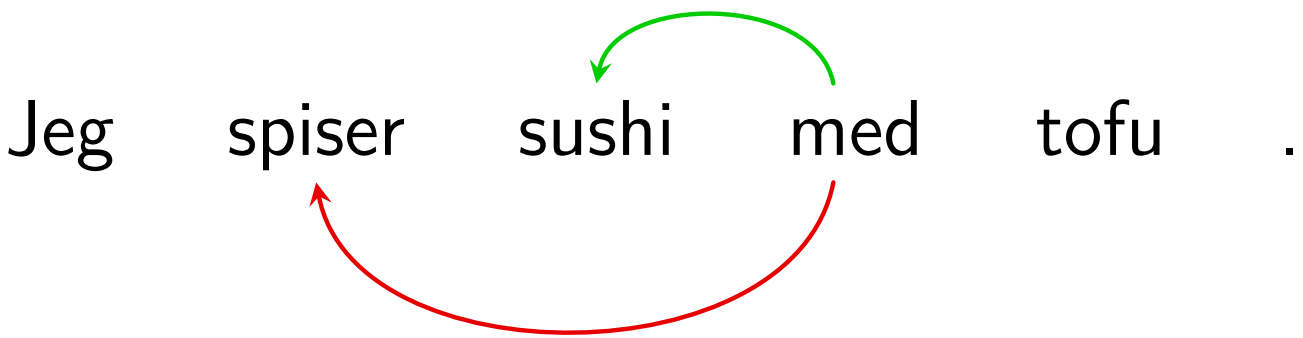
Flertydighet på setningsnivå





Flertydighet på setningsnivå





Entydiggjøring

- Vi mennesker tolker språklige uttrykk basert på delt bakgrunnskunnskap og gjensidige forventninger i en gitt kontekst.
- Språkforståelse handler mye om entydiggjøring.
- Språkteknologi, og IN1140, handler i stor grad om strategier for hvordan maskiner kan takle dette.

Historisk: To paradigmer

- Regel-baserte metoder
- Data-drevne metoder / maskinlæring
- Følg instrukser vs. lær fra erfaring

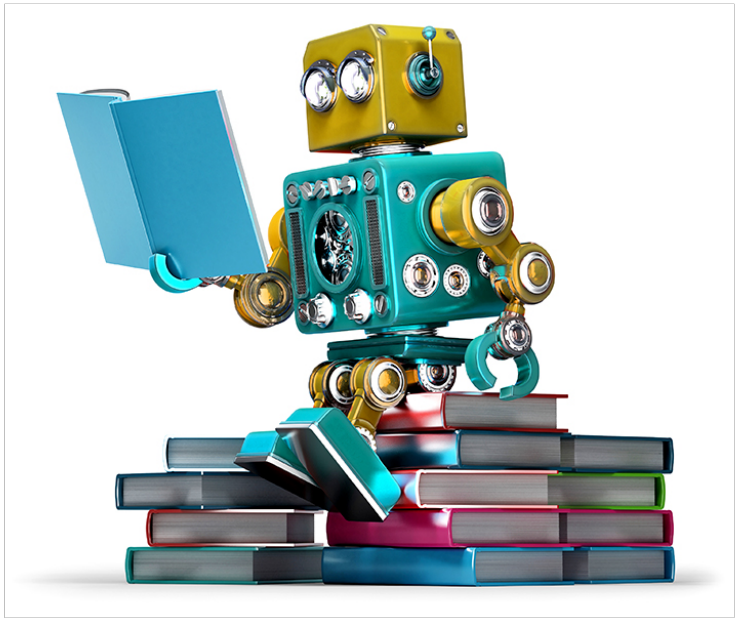
Regel-baserte metoder

- NLP/AI på 1950–90-tallet
- 'Symbolske metoder'
- Formelle regler for å analysere språk og beskrive verden.
- F.eks. manuelt utformede regler og leksikon
- Manuell innsats fra en ekspert.
- Problem: skalerer ikke.

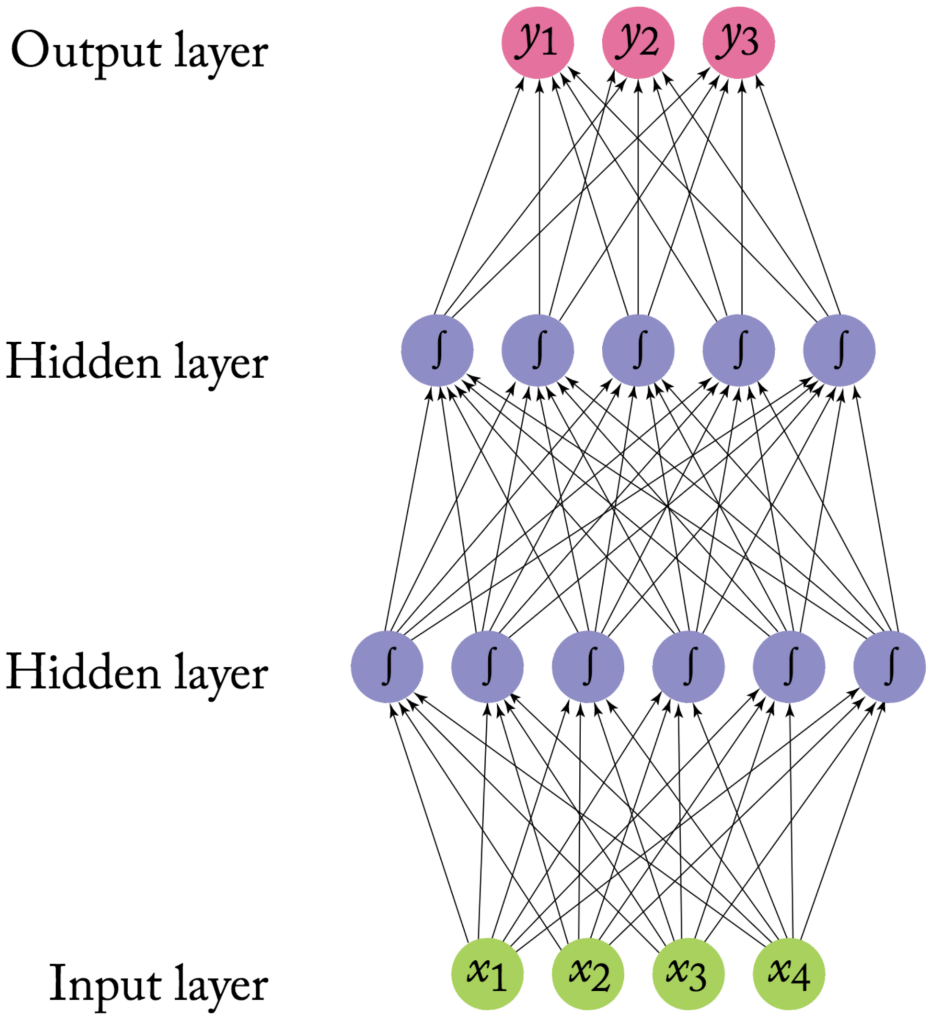


Data-drevne metoder

- NLP/AI på 2000-tallet
- 'Empiriske metoder'
- Maskinlæring: algoritmer som selv lærer fra data.
- I stedet for manuell koding av kunnskap:
- Finn statistiske mønstre i observert språkbruk.



Et andre paradigmeskifte



- NLP/AI fra 2011/2015–
- Kunstige nevrale nett
- En familie kraftige maskinlæringmodeller
- Har revolusjonert mange felt, inkludert språkteknologi
- 'Dyp Læring': NN med flere skjulte lag.

IN1140 Praktiske detaljer



IN1140 og tverrfaglighet

- Stoffet vi dekker i IN1140 tar også for seg stoff fra flere ulike felt.
- Innføring i lingvistikk,
- grunnleggende sannsynlighetsregning,
- programmering, og
- språkteknologiske anvendelser.

IN1140 og tverrfaglighet

- Stoffet vi dekker i IN1140 tar også for seg stoff fra flere ulike felt.
- Innføring i lingvistikk,
- grunnleggende sannsynlighetsregning,
- programmering, og
- språkteknologiske anvendelser.

- Gjør deg godt rustet for flere viderekommende BSc-emner, f.eks
  - Flere emner på nytt Bachelorprogram in ML & KI
  - IN3120 – Søketeknologi

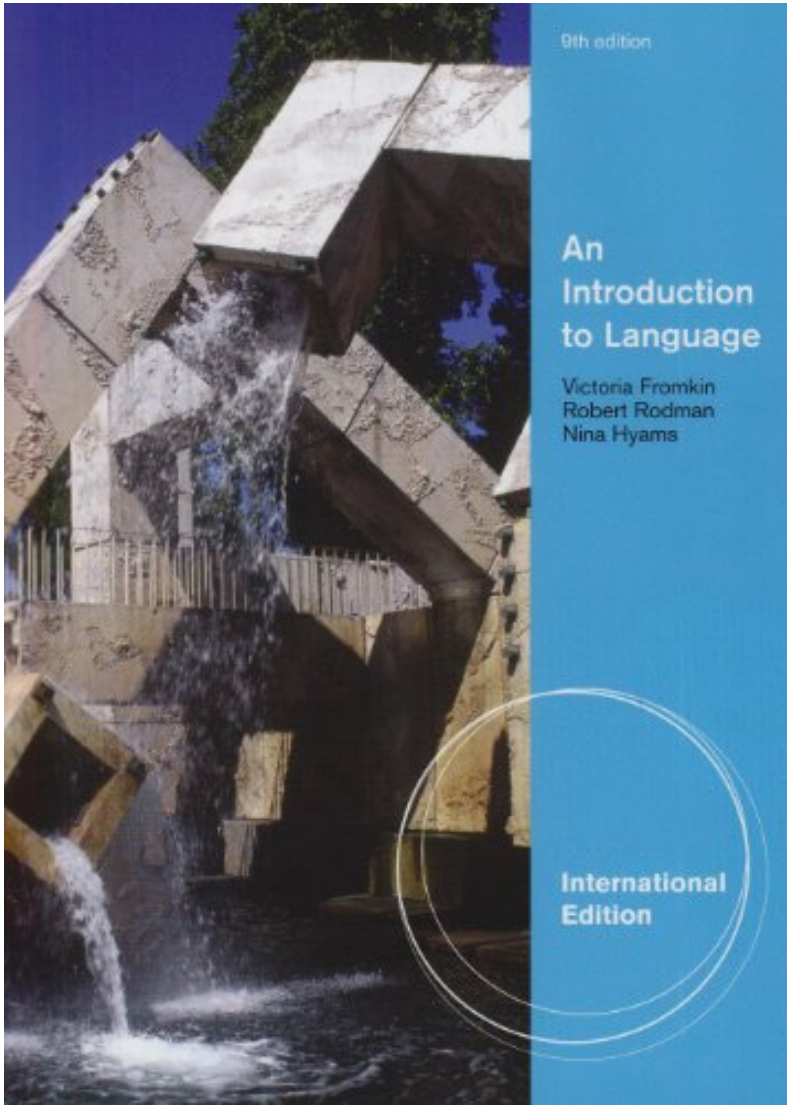
IN1140 og tverrfaglighet

- Stoffet vi dekker i IN1140 tar også for seg stoff fra flere ulike felt.
- Innføring i lingvistikk,
- grunnleggende sannsynlighetsregning,
- programmering, og
- språkteknologiske anvendelser.

- Gjør deg godt rustet for flere viderekommende BSc-emner, f.eks
  - Flere emner på nytt Bachelorprogram in ML & KI
  - IN3120 – Søketeknologi

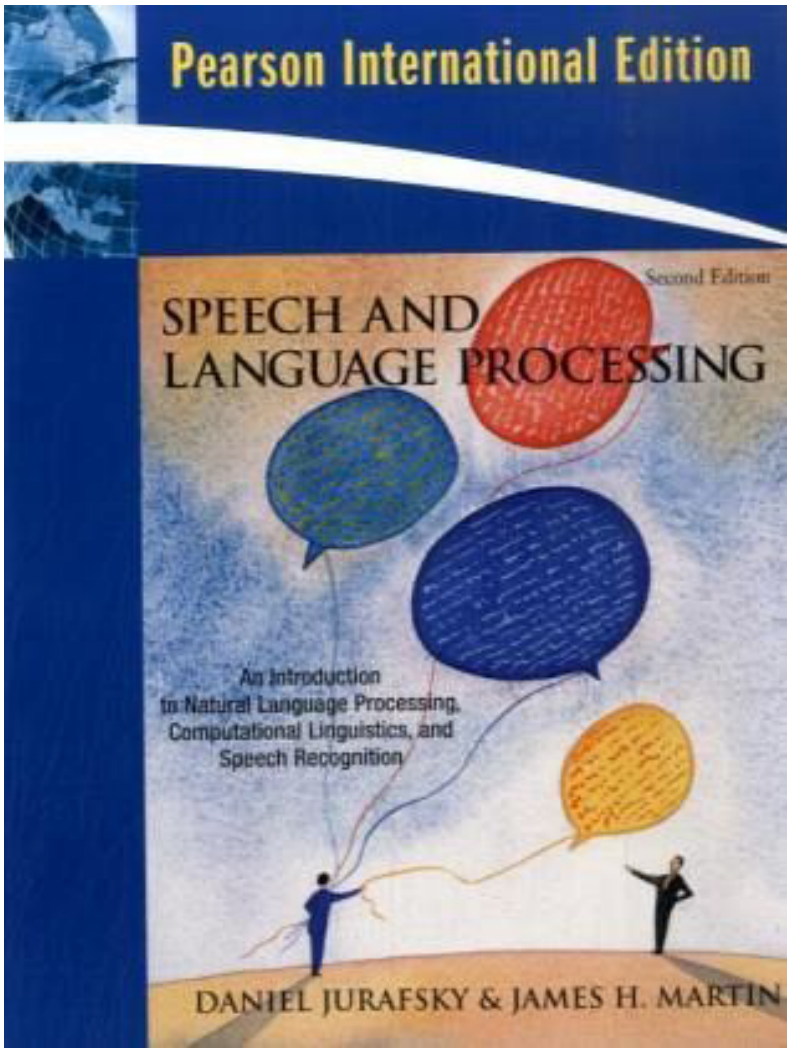
- Mulig å fortsette på MSc-nivå, f.eks:
  - IN4080 – Natural Language Processing
  - IN5550 – Neural Methods in NLP

Pensumlitteratur



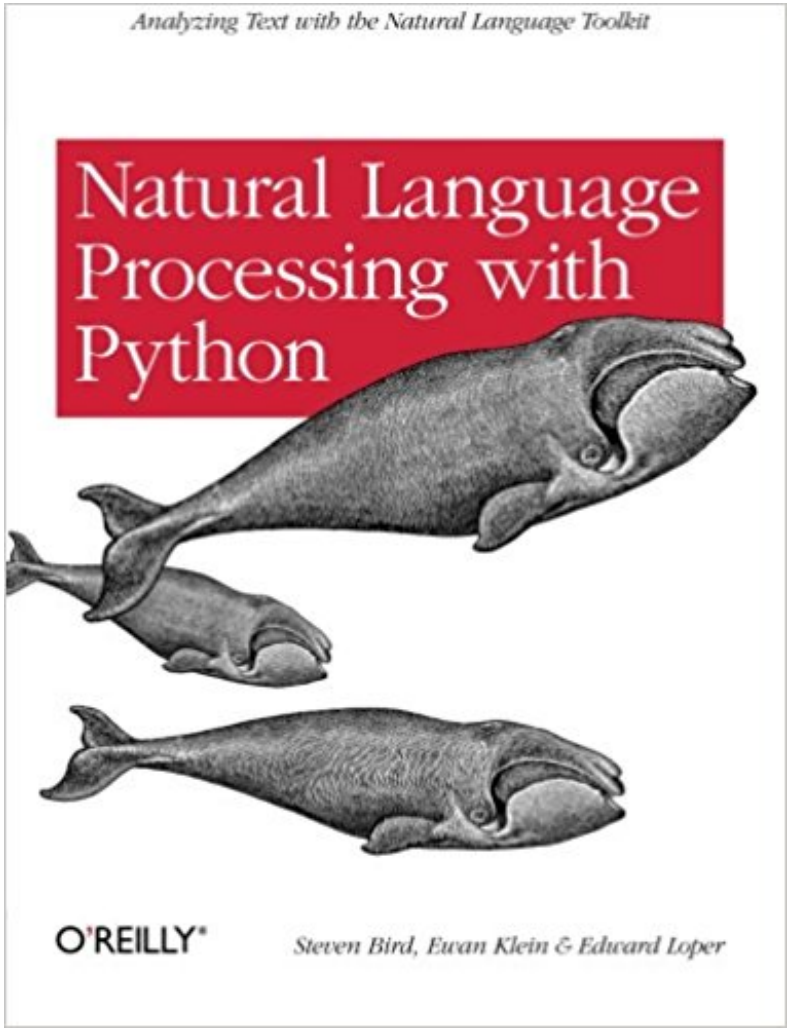
- *An Introduction to Language* av Fromkin, Rodman & Hyams
- Utvalgte deler (ca 5 kapitler)

Pensumlitteratur



- *Speech and Language Processing* av Jurafsky & Martin
- Utvalgte deler
- Gratis nettbok:
  `https://web.stanford.edu/~jurafsky/slp3/`

Pensumlitteratur



- *Natural Language Processing with Python*,
  av Bird, Klein & Loper
- Oppdatert for Python 3 og NLTK 3 (Natural Language Toolkit)
- Utvalgte deler
- Gratis nettbok:
  `http://www.nltk.org/book/`

Python

- Progammering lærer dere først og fremst i IN1000, ikke IN1140.
- Forelesningene i IN1140 kommer til fokusere på teori.
- Samtidig ønsker vi å implementere stoffet i praksis, i Python.
- Implementasjon blir fokus på gruppene og innleveringene.
- Kræsjkurs i Python-programmering på de første gruppetimene.
- Viktig med en del egeninnsats i starten for å henge med.



Forelesninger og grupper

Tid & sted

- Forelesninger: opptak fra tidligere år (med unntak av første og siste forelesning)
- Stor grad av selvstudium
- Gruppe 1: tors. 10:15–12:00, Datastue Chill (OJD/IFI)
- **NB!** Første gruppetime er neste uke.

Spørsmål og hjelp

- Gruppetimene: Gruppelærerene er der for å hjelpe og veilede.
- Diskusjonsforum:
  `https://astro-discourse.uio.no/c/in1140-25h`
- `in1140-hjelp [at] ifi.uio.no`: Felles adresse til fag-/gruppelærere.
- Send alltid henvendelser til lista framfor individuelle adresser.
- Faglige spørsmål skal rettes til forumet.



Beskjeder

- Husk å sjekke UiO-eposten din og beskjedlisten på semestersiden.
- `http://www.uio.no/studier/emner/matnat/ifi/IN1140/h25/`



Obligatoriske innleveringer

- 2 obliger: gruppearbeid (2–3 studenter)
- Oblig 1 har to deler (a + b).
- Oblig 2 har to deler (a + b).
- Dvs. 4 innleveringer tilsammen: 1a + 1b, 2a + 2b.
- Gruppebytte mulig mellom oblig 1 og 2, men ikke mellom a/b-innleveringer.
- Begge obligene må bestås for å kunne ta eksamen.
- Ingen omlevering.

Obligatoriske innleveringer

- 2 obliger: gruppearbeid (2–3 studenter)
- Oblig 1 har to deler (a + b).
- Oblig 2 har to deler (a + b).
- Dvs. 4 innleveringer tilsammen: 1a + 1b, 2a + 2b.
- Gruppebytte mulig mellom oblig 1 og 2, men ikke mellom a/b-innleveringer.
- Begge obligene må bestås for å kunne ta eksamen.
- Ingen omlevering.

Poengsystemet

- Man kan oppnå opptil 100 poeng per innlevering
- For å bestå en oblig kreves minst 120 poeng (av 200 mulige; 60%).
- Eksempel:
  - 45 poeng på 1a
  - 80 poeng på 1b
  - = 125 poeng på oblig 1 (= bestått).

Mer om obligene

- Absolutte frister:
- Utsettes *kun* ved egenmelding (opptil 3 dager) eller legeerklæring.
- Kopiering/plagiat godtas ikke. Sett deg inn i reglene.
- Husk at hvis du distribuerer løsningsforslaget ditt på nett (f.eks via Github), kan du bidra til juks. Styr unna.
- Benytt deg av gruppeundervisningen, og planlegg tiden din.
- Konkurranse: den/de som får flest poeng tilsammen på obligene gjennom semesteret får en premie (overraskelse)!

Eksamen

- Skriftlig (digital) skoleeksamen på fire timer
  - 17 desember kl. 15:00
- Pensumlitteratur + forelesningsnotater
- NB! Ikke en programmeringseksamen.
- Fokus på teoretiske konsepter.

Suksessoppskrift

- Emnesiden: timeplan, pensum, lesehenvisninger, beskjeder etc.
- Lesehenvisninger: selvstudier i tillegg til forelesninger/grupper
- Still spørsmål
- Gå på gruppetimene:
  - forbered deg
  - delta aktivt
  - gjør oppgaver (også de ikke-obligatoriske!)
- Benytt deg av medstudentene dine

Neste uke

- Hva lærer jeg i IN1140?
- Hva er lingvistikk?
- Språkteknologiske komponenter
- Metoder