

INF1820: Introduksjon til språk-og kommunikasjonsteknologi

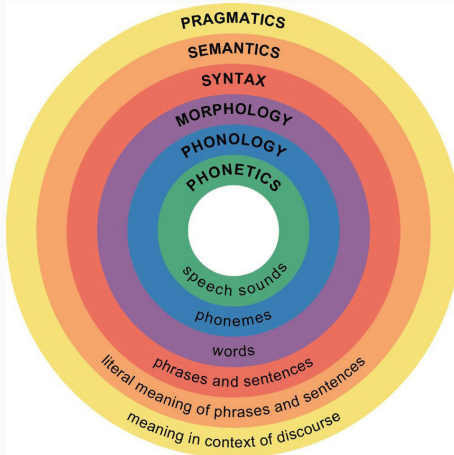
Ellevte forelesning

Lilja Øvrelid

24 april, 2017

Semantisk språkteknologi

Lingvistiske nivåer



- studiet av betydning slik det uttrykkes gjennom språk
- betydning til morfemer, ord, fraser og setninger
 - leksikal semantikk
 - setningssemantikk
 - (pragmatikk: hvordan konteksten påvirker betydning)

- hva slags oppgaver inngår i semantisk analyse?
- med hva slags metoder kan disse oppgavene løses?

Ordbetydning ("word sense")

Flertydighet

- *The astronomer married the star*
- *You are free to execute your laws, and your citizens, as you see fit* (Star Trek, Next Generation)
- *Oh, flowers are common here, Miss Fairfax, as people are in London* (Oscar Wilde, The Importance of Being Earnest)

Hva slags applikasjoner trenger kunnskap om ordbetydning?

- *The dog bit the mailman* er ikke det samme som *The mailman bit the dog*
- sammenligne med *The dog was bitten by the mailman*
- semantiske roller beskriver "hvem som gjør hva mot hvem"
[The mailman]_{AGENT} bit [the dog]_{PATIENT}

- *Apple bought Cisco*
- *Apple acquired Cisco*
- *Cisco was taken over by Apple*

hva slags applikasjoner kan bruke denne informasjonen?

Semantiske relasjoner mellom ord

- synonymer (samme betydning)
 - *ascend/rise*
 - *sweater/pullover*
- antonymer (motsetninger)
 - *good/evil*
 - *ascend/descend*
- hypernym/hyponym (mer generell, mer spesifikk)
 - *tree/birch*
 - *animate object / mammal / whale*
- meronymi (del-helhet)
 - *knob/door*
 - *wheel/car*

- *The chairman announced yesterday that they would have the problem solved within three days*
 - når ble problemet løst dersom setningen forekom i en avis 10 mai, 2007?
- *Book me a flight on the 7:35 tomorrow*

- spørsmål: *Who bought Cisco?*
- forventet svar: *Apple bought Cisco*
 - *Cisco's acquisition by Apple* → (entails) *Apple bought Cisco*

- *Apple acquired Cisco*
- *Apple did not acquire Cisco*
- *Apple failed to acquire Cisco*
- *Apple denied not acquiring Cisco*

- automatisk negasjonsanalyse temaet for en internasjonal forskningskonkurranse
 - *SEM Shared Task on Negation Resolution
 - system som angir
 - negation **cue**
 - negation scope
 - negated *event*
 - There was **no** answer.

- 9 lag fra hele verden deltok
- UiO (LTG) deltok med to systemer:
 - UiO_1 : rangering av syntaktiske “stier”
 - UiO_2 : sekvensiell klassifisering
- UiO kom på første, andre og tredje plass

- MSc-oppgave IFI (2016): *A portable toolkit for detecting negation*
- Tilgjengelig her: <https://www.duo.uio.no/handle/10852/54815?show=full>
- Nylig presentert på internasjonal konferanse (EACL, Valencia)

- *X announced that Apple acquired Cisco*
- *X denied that Apple acquired Cisco*
- *Apple may possibly have acquired Cisco*
- *It is likely that Apple acquired Cisco*

- klassifisering
 - gitt et ord i en setning, og en liste av mulige betydninger, velg en betydning
 - gitt et predikat i en setning, finn dets semantiske roller
- automatisk tilegnelse av semantisk informasjon fra rå tekst
 - hvilke ord og fraser betyr det samme
 - distribusjonell semantikk:
<http://vectors.nlpl.eu/explore/embeddings/en/>

⇒ **Maskinlæring**

- 2000-tallet →: empirisk revolusjon
- **Maskinlæring**
 - Datamaskiner kan lære fra data: fange opp mønstre og generalisere til nye eksempler



Semantiske ressenser

- klassifisering forutsetter treningsdata
- leksikalske databaser (WordNet, FrameNet)
- korpuser annotert med semantisk informasjon (PropBank)

- manuelt konstruert database
- betydningen til ord karakteriseres gjennom **relasjoner** til andre ord
- semantiske konsepter karakteriseres gjennom relasjoner til andre konsepter
- hva slags relasjoner kan det være snakk om?

- mellom ord:
 - synonymi (samme betydning)
 - synonymi-relasjonen grupperer ord i synonymmengder, såkalte **synsets**
- mellom konsepter (=synsets)
 - hypernyymi (mer generell, mer spesifikk)
 - varierer noe, men antonymi og meronymi er også spesifisert for noen synsets

- elektronisk leksikon
 - online grensesnitt
 - lastes ned
 - tilgjengelig på <http://wordnet.princeton.edu/>
 - også tilgjengelig via NLTK
- består av tre separate databaser:
 1. substantiv (117097 lemmaer)
 2. verb (11488 lemmaer)
 3. adjektiv og adverb (22141 adjektiver, 4601 adverb)

- verbet *skim*: synonymer, definisjoner og eksempler

Verb

- **S:** (v) [plane](#), **skim** (travel on the surface of water)
- **S:** (v) [skim over](#), **skim** (move or pass swiftly and lightly over the surface of)
- **S:** (v) [scan](#), **skim**, [rake](#), [glance over](#), [run down](#) (examine hastily) "*She scanned the newspaper headlines*"
- **S:** (v) **skim**, [skip](#), [skitter](#) (cause to skip over a surface) "*Skip a stone across the pond*"
- **S:** (v) **skim** (coat (a liquid) with a layer)
- **S:** (v) **skim**, [skim off](#), [cream off](#), [cream](#) (remove from the surface) "*skim cream from the surface of milk*"
- **S:** (v) **skim**, [skim over](#) (read superficially)

- synsets er koblet sammen ved
 - hyponym/hypernym relasjonen (hovedhierarkiet)
 - meronymi: del-helhet relasjoner
 - komponent/del (*table – leg, finger – hand, petal – blossom*)
 - medlem av en gruppe *forest – tree, student – class*
 - materiale et objekt er laget av (*oxygen – water*)
- ord er koblet sammen ved antonymi

WordNet: substantiv eksempel

- **S: (n) cat, true cat** (feline mammal usually having thick soft fur and no ability to roar: domestic cats; wildcats)
 - [direct hyponym](#) / [full hyponym](#)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (n) feline, felid** (any of various lithe-bodied roundheaded fissioned mammals, many with retractile claws)
 - **S: (n) carnivore** (a terrestrial or aquatic flesh-eating mammal) "*terrestrial carnivores have four or five clawed digits on each limb*"
 - **S: (n) placental, placental mammal, eutherian, eutherian mammal** (mammals having a placenta; all mammals except monotremes and marsupials)
 - **S: (n) mammal, mammalian** (any warm-blooded vertebrate having the skin more or less covered with hair; young are born alive except for the small marsupials)
 - **S: (n) vertebrate, craniate** (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull)
 - **S: (n) chordate** (any animal of the phylum Chordata having a notochord or spinal column)
 - **S: (n) animal, animate being, beast, brute, creature, fauna** (a living organism characterized by voluntary movement)
 - **S: (n) organism, being** (a living thing that has (or can develop) the ability to act or function independently)
 - **S: (n) living thing, animate thing** (a living (or once living) entity)
 - **S: (n) whole, unit** (an assemblage of parts that is regarded as a single entity) "*how big is that part compared to the whole*"
 - **S: (n) object, physical object** (a tangible and visible entity; an entity that can cast a shadow) "*it was full of objects*"
 - **S: (n) physical entity** (an entity that has physical existence)
 - **S: (n) entity** (that which is perceived or known or inferred to have its own distinct existence (I have a feeling that she is still alive))
- **S: (n) boy, cat, bambos, boyo** (an informal term for a youth or man) "*a nice boy*"; "*the boy's only doing it for some doll*"

WordNet: substantiv eksempler hyponymi, meronymi, holonymi

- **S: (n) mammal, mammalian** (any warm-blooded vertebrate having the skin more or less covered with hair; young are born alive except for monotremes)
 - *direct hyponym / full hyponym*
 - **S: (n) female mammal** (animals that nourish their young with milk)
 - **S: (n) tusker** (any mammal with prominent tusks (especially an elephant or wild boar))
 - **S: (n) prototherian** (primitive oviparous mammals found only in Australia and Tasmania and New Guinea)
 - **S: (n) metatherian** (primitive pouched mammals found mainly in Australia and the Americas)
 - **S: (n) placental, placental mammal, eutherian, eutherian mammal** (mammals having a placenta; all mammals except monotremes)
 - **S: (n) fossorial mammal** (a burrowing mammal having limbs adapted for digging)
 - *part meronym*
 - **S: (n) coat, pelage** (growth of hair or wool or fur covering the body of an animal)
 - **S: (n) hair, pilus** (any of the cylindrical filaments characteristically growing from the epidermis of a mammal) "there is a hair on my arm"
 - *member holonym*
 - **S: (n) Mammalia, class Mammalia** (warm-blooded vertebrates characterized by mammary glands in the female)
 - *domain term category*

- semantikk for hendelser ('events', verb) er ganske forskjellig fra semantikk for entiteter (substantiver)
- relasjoner
 - troponymi:
 - *A verb expressing a specific manner of elaboration of another verb. X is a troponym of Y if to X is to Y in some manner*
 - *wade – walk*
 - entailment
 - *A verb X entails Y if X cannot be done unless Y is, or has been, done*
 - *walking – stepping*

WordNet: verb eksempel

- **S: (v) jump, leap, bound, spring** (move forward by leaps and bounds) "*The horse bounded across the meadow*"; "*The child leapt across the puddle*"
 - *direct troponym* / *full troponym*
 - **S: (v) pronk** (jump straight up) "*kangaroos pronk*"
 - **S: (v) bounce, resile, take a hop, spring, bound, rebound, recoil, reverberate, ricochet** (spring back; spring away from an impact) "*The ball bounced off the wall*"
 - **S: (v) burst** (move suddenly, energetically, or violently) "*He burst out of the house into the cool night*"
 - **S: (v) bounce** (leap suddenly) "*He bounced to his feet*"
 - **S: (v) capriole** (perform a capriole, of horses in dressage)
 - **S: (v) galumph** (move around heavily and clumsily) "*the giant tortoises galumphed around in their pen*"
 - **S: (v) ski jump** (jump on skis)
 - **S: (v) saltate** (leap or skip, often in dancing) "*These fish swim with a saltating motion*"
 - **S: (v) vault** (bound vigorously)
 - **S: (v) leapfrog** (jump across) "*He leapfrogged his classmates*"
 - **S: (v) vault, overleap** (jump across or leap over (an obstacle))
 - **S: (v) curvet** (perform a leap where both hind legs come off the ground, of a horse)
 - **S: (v) hop, skip, hop-skip** (jump lightly)
 - **S: (v) caper** (jump about playfully)
 - **S: (v) hop** (make a jump forward or upward)
 - *direct hypernym* / *inherited hypernym* / *sister term*
 - **S: (v) move** (move so as to change position, perform a nontranslational motion) "*He moved his hand slightly to the right*"

- S: (v) **walk** (use one's feet to advance; advance by steps) "*Walk, don't run!*"; "*We walk*"
 - direct troponym / full troponym
 - verb group
 - direct hypernym / inherited hypernym / sister term
 - entailment
 - S: (v) step (shift or move by taking a step) "*step back*"

- **deskriptive** adjektiver
 - organisert i grupperinger basert på likhet ("similarity"), relatert via antonymi
- **relasjonelle** adjektiver, linket til sine substantiver
 - *nasal – nose*
- liten gruppe adjektiver "reference-modifying"
 - *former, alleged*

WordNet: adjektiv eksempel

- **S: (adj) cheap, inexpensive** (relatively low in price or charging low prices) "*it would have been cheap at twice the price*";
 - *similar to*
 - **S: (adj) bargain-priced, cut-rate, cut-price** (costing less than standard price) "*buying bargain-priced clothes for*"
 - **S: (adj) catchpenny** (designed to sell quickly without concern for quality) "*catchpenny ornaments*"
 - **S: (adj) dirt cheap** (very cheap) "*a dirt cheap property*"
 - **S: (adj) low-budget** (made on or suited to a limited budget) "*a low-budget movie*"; "*a low-budget menu*"
 - **S: (adj) low-cost, low-priced, affordable** (that you have the financial means for) "*low-cost housing*"
 - **S: (adj) nickel-and-dime** (low-paying) "*a nickel-and-dime job*"
 - **S: (adj) sixpenny, threepenny, twopenny, tuppenny, two-a-penny, twopenny-halfpenny** (of trifling worth)
 - *antonym*
 - **W: (adj) expensive** [Opposed to: **cheap**] (high in price or charging high prices) "*expensive clothes*"; "*an expensive*"
 - *derivationally related form*

- hvor mange betydninger har et ord?
 - antall synsets ordet forekommer i
- nærhet i betydning kan utledes fra nærhet i hierarkiet
 - korteste stien via hyponym/hyponym-linkene mellom synsets

- utgangspunkt for Word Sense Disambiguation
 - merke forekomster av et ord med riktig betydning (=synset)
- generaliserer over ord via hypernyymi-relasjonen
 - fra *cat* til *living being*
- generalisere over synonymer
- ...

- aspekt ved setningsbetydning: hvilke roller de forskjellige deltagerene inntar
 - *Gina* hevet *bilen* med *jekken*
 - *Gina* – deltageren som er ansvarlig for å utføre handlingen beskrevet av verbet
 - *bilen* – blir påvirket av handlingen
 - *jekken* – middelet som Gina bruker til å utføre handlingen
- semantiske roller beskriver den semantiske relasjonen som argumenter har til handlingen beskrevet av verbet

- eksempelet:

Gina *hevet* *bilen* *med* *jekken*
AGENT THEME INSTRUMENT

Problemer med semantiske roller

- ikke full enighet rundt rolleinventaret
- vanskelig å formulere formelle definisjoner av roller
- \Rightarrow generaliserte semantiske roller
 - PROTO-AGENT, PROTO-PATIENT
- verbspesifikke roller
- semantiske ressurser med informasjon om semantiske roller:
PropBank og **FrameNet**

- inneholder alle setningene i Penn Treebank
- roller er (stort sett) verbspesifikke
 - Arg0, Arg1 = PROTO-AGENT, PROTO-PATIENT
 - Arg2 ... verbspesifikke

agree.01

Arg0 Agreeer

Arg1 Proposition

Arg2 Other entity agreeing

Ex1 [Arg0 *The group*] *agreed* [Arg1 *it wouldn't make an offer unless ...*]

Ex2 [ArgM-TMP *Usually*] [Arg0 *John*] *agrees* [Arg2 *with Mary*] [Arg1 *on everything*]

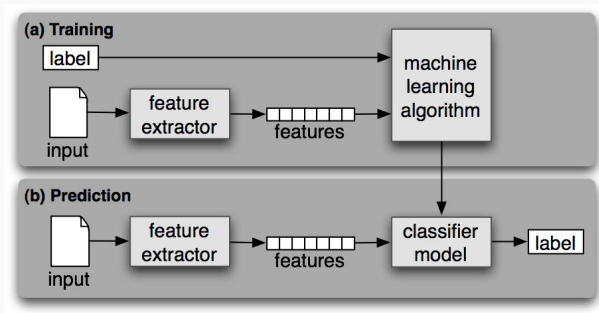
- applikasjon: Semantic Role Labeling
- gitt et predikat i en setning, finn dets semantiske roller
- gir oss en felles representasjon for:
 - [Arg_0 Big Fruit Co.] increased [Arg_1 the price of bananas]
 - [Arg_1 The price of bananas] was increased again by [Arg_0 Big Fruit Co.]
 - [Arg_1 The price of bananas] increased [Arg_2 5%]

Semantisk klassifisering (WSD)

- Word Sense Disambiguation (WSD) – aktivt felt innenfor språkteknologi
 - gitt en setning med et spesifikt målord ("target word") og en liste med betydninger (f.eks. fra WordNet)
 - angi korrekt betydning for målordet i den setningen
- klassifisering basert på et annotert datasett

- sentral metode innenfor maskinlæring
- automatisk avgjøre hvilken kategori en observasjon tilhører
- basert på **treningsdata**: observasjoner der kategorien er kjent
 - e-post \rightarrow {spam, ikke-spam}
 - pasient \rightarrow diagnose
- **supervised** klassifisering: klassifisering som benytter treningsdata

Statistisk klassifisering



Betydningsdisambiguering

- første skritt består i å hente ut trekk (“features”) fra treningsdataene
- eksempel: setninger merket med betydning
 - **SKIM** the pages for a clearer insight: [Reading](#)
 - She **SKIMS** through the novel which seems to fascinate them: [Reading](#)
 - Remove the vanilla pod, **SKIM** the jam, and let it cool: [Removing](#)
 - We **SKIMMED** across the surface of that sodding lake whilst all around us gathered the dark hosts of hell: [Self_Motion](#)
- hvilke trekk (“features”) kan vi bruke for å skille mellom de forskjellige betydningene?

Trekk: bag-of-words

- **SKIM** the pages for a clearer insight: [Reading](#)
- She **SKIMS** through the novel which seems to fascinate them: [Reading](#)
- Remove the vanilla pod, **SKIM** the jam, and let it cool: [Removing](#)
- We **SKIMMED** across the surface of that sodding lake whilst all around us gathered the dark hosts of hell: [Self_Motion](#)

henter ut alle ord (**ikke** ordnet):

- a, clearer, for, insight, pages, the: [Reading](#)
- fascinate, novel, seems, she, the, them, through, to, which: [Reading](#)
- and, cool, it, jam, let, pod, remove, the, the, vanilla: [Removing](#)
- across, all, around, dark, gathered, hell, hosts, lake, of, of, sodding, surface, that, the, the, us, we, whilst [Self_Motion](#)

Trekk: bag-of-words (forts.)

- bred kontekst: (20 ord, 50 ord på hver side av målordet)
modell av generelt tema/domene
- smal kontekst (1-5 ord på hver side)
kollokasjoner, grunn modell av syntaktisk kontekst

- konteksten til målordet kan representeres ved
 - ordformer
 - lemmaer
 - ordklassetagger
 - kombinasjon av disse
 - $[w_{i-2}, POS_{i-2}, w_{i-1}, POS_{i-1}, w_{i+1}, POS_{i+1}, w_{i+2}, POS_{i+2}]$
 - Remove the vanilla pod, SKIM the jam, and let it cool:
Removing
 - **trekkvektor**: [vanilla, JJ, pod, NN, the, DT, jam, NN]

Trekk: syntaktiske funksjoner

- SKIM the pages for a clearer insight: [Reading](#)
 - She SKIMS through the novel which seems to fascinate them: [Reading](#)
 - Remove the vanilla pod, SKIM the jam, and let it cool: [Removing](#)
 - We SKIMMED across the surface of that sodding lake whilst all around us gathered the dark hosts of hell: [Self_Motion](#)
-
- verbets argumenter:
 - direkte_objekt: [Reading](#)
 - subjekt, pp_through: [Reading](#)
 - direkte_objekt: [Removing](#)
 - subjekt, pp_across: [Self_Motion](#)
 - kombinasjoner (argumenters hovedord)
 - direkte_objekt / pages: [Reading](#)
 - subjekt / she, pp_through / novel: [Reading](#)
 - direkte_objekt / jam: [Removing](#)
 - subjekt / we, pp_across / surface: [Self_Motion](#)

Trekk: n-gram av ord

- n-gram av ord i nærheten av målordet
 - $n=1,2,3$
 - kan også bruke ordformer, lemmaer, ordklasser
- eksempel:
 - **SKIM** the pages for a clearer insight: [Reading](#)
 - She **SKIMS** through the novel which seems to fascinate them: [Reading](#)
 - Remove the vanilla pod, **SKIM** the jam, and let it cool: [Removing](#)
 - We **SKIMMED** across the surface of that sodding lake whilst all around us gathered the dark hosts of hell: [Self_Motion](#)
- trigram:
 - -, -, -, the, pages, for
 - -, -, She, through, the, novel
 - the, vanilla, pod, the, jam, and
 - -, -, We, across, the, surface

- gitt treningsdataene og trekkvektorene, kan en rekke forskjellige maskinlæringsalgoritmer brukes til å trene en klassifiserer
- her skal vi se på **Naive Bayes**-klassifisering
- bruker informasjon om ord i konteksten for disambiguering av betydning
- enkel metode, mye brukt i WSD

- hovedantagelse: for å finne en betydning \hat{s} hentet fra alle mulige betydninger av et ord S for en trekkvektor \vec{f} må vi beregne den mest sannsynlige betydningen, gitt vektoren

$$\hat{s} = \operatorname{argmax}_{s \in S} P(s | \vec{f})$$

- men problematisk å trene direkte: "sparse data"-problemet
- kan bruke **Bayes teorem** her også!

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Naive Bayes

- omformulering ved Bayes teorem:

$$\hat{s} = \operatorname{argmax}_{s \in S} \frac{P(\vec{f} | s)P(s)}{P(\vec{f})}$$

- men fremdeles mangelfulle data for denne beregningen
- bryter opp trekkvektoren og ser på individuelle trekk i kombinasjon med betydninger
- en **uavhengighetsantagelse**: trekk er uavhengige av andre trekk

$$P(\vec{f} | s) \approx \prod_{j=1}^n P(f_j | s)$$

- hvilke andre uavhengighetsantagelser har vi sett?

- Naive Bayes klassifiserer

$$\hat{s} = \operatorname{argmax}_{s \in \mathcal{S}} P(s) \prod_{j=1}^n P(f_j | s)$$

- vi **trener** klassifisereren ved å beregne sannsynligheter fra et korpus (MLE)
- 2 sannsynligheter:

1. prior-sannsynligheten for betydningen $P(s)$

$$P(s_i) = \frac{\text{count}(s_i, w_j)}{\text{count}(w_j)}$$

2. sannsynligheten for individuelle trekk $P(f_j | s)$

$$P(f_j | s) = \frac{\text{count}(f_j, s)}{\text{count}(s)}$$

trekkvektorer (bag-of-words), der siste element angir betydningsklassen

- [a, clearer, for, insight, pages, the, [Reading](#)]
- [fascinate, novel, seems, she, the, them, through, to, which, [Reading](#)]
- [and, cool, it, jam, let, pod, remove, the, the, vanilla, [Removing](#)]
- [across, all, around, dark, gathered,hell, hosts, lake, of, of, sodding, surface, that, the, the, us, we, whilst, [Self_Motion](#)]

1. prior-sannsynligheten for betydningen $P(\textit{Reading})$

- $P(s_i) = \frac{\textit{count}(s_i, w_j)}{\textit{count}(w_j)}$
- $P(\textit{Reading}) = \frac{\textit{count}(\textit{Reading}, \textit{skim})}{\textit{count}(\textit{skim})}$
- $P(\textit{Reading}) = \frac{2}{4} = 0.5$

Eksempel forts.

- [a, clearer, for, insight, pages, the, Reading]
- [fascinate, novel, seems, she, the, them, through, to, which, Reading]
- [and, cool, it, jam, let, pod, remove, the, the, vanilla, Removing]
- [across, all, around, dark, gathered, hell, hosts, lake, of, of, sodding, surface, that, the, the, us, we, whilst, Self_Motion]

2. sannsynligheten for individuelle trekk $P(f_j|s)$

- $P(f_j|s) = \frac{\text{count}(f_j,s)}{\text{count}(s)}$
- $P(\text{pages}|\text{Reading}) = \frac{1}{2} = 0.5$
- $P(\text{pages}|\text{Removing}) = \frac{0}{1} = 0$
- $P(\text{pages}|\text{Self_Motion}) = \frac{0}{1} = 0$
- ...
- $P(\text{jam}|\text{Reading}) = \frac{0}{2} = 0$
- $P(\text{jam}|\text{Removing}) = \frac{1}{1} = 1.0$
- $P(\text{jam}|\text{Self_Motion}) = \frac{0}{1} = 0$
- ...

Eksempel forts.

- vi har nå en NB-modell som vi kan bruke til å klassifisere en ny og usett setning:
 - I like to SKIM through the novel
 - [I, like, novel, the, through, to, ??]
- vi bruker den velkjente formelen:

$$P(s) \prod_{j=1}^n P(f_j|s)$$

- itererer gjennom alle betydningene og trekkene, ganger sammen
- og velger den betydningen (s) som gir høyest sannsynlighet

- Syntaks:
 - skrive en liten grammatikk som analyserer en (del)mengde norske setninger
 - teste med nltk (RecursiveDescent-parseren)
- Semantikk:
 - annotering av ordbetydning med bruk av WordNet
 - WSD med en Naive Bayes-klassifiserer (lite treningssett + testsett)

Oppsummering

- Rekke oppgaver inngår i semantisk analyse
 - ordbetydningsdisambiguering (WSD)
 - semantiske roller
 - parafrasering
 - temporal analyse
 - entailment
 - negasjon
 - ...
- Sentral metode: [klassifisering](#)

- For klassifisering trenger vi treningsdata
- Semantiske ressurser
 - WordNet
 - leksikal database
 - innholdsord: substantiver, verb, adjektiver
 - bygget rundt leksikale relasjoner som synonymi, hyponymi, meronymi, etc.
 - PropBank/FrameNet
 - forskjellige ressurser for semantiske roller
 - korpus vs database
 - verbspesifikk vs ramme ("frame")
 - ...

- Nærmere kikk på betydningsdisambiguering
- Trekkrepresentasjon av treningsdata
 - ord
 - lemma
 - ordklasse
 - syntaktisk funksjon
 - etc.
- Naive Bayes-klassifisering
 - hvordan vi kan beregne den mest sannsynlige betydningen for et ord:

$$\hat{s} = \operatorname{argmax}_{s \in S} P(s | \vec{f})$$