

IN1140: Introduksjon til språkteknologi

Forelesning #09

Lilja Øvrelid

Universitetet i Oslo

12 oktober 2020



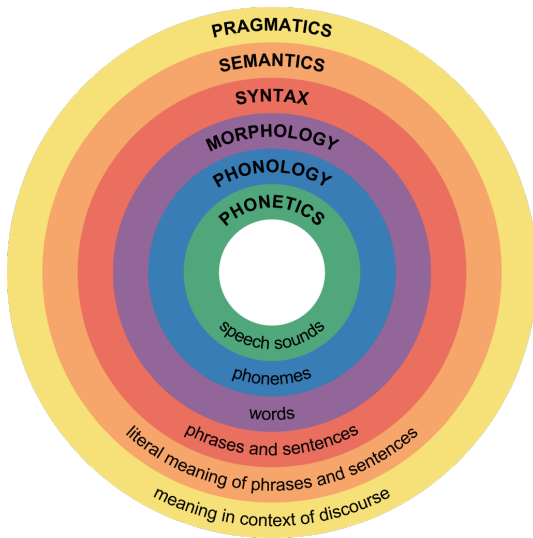
De forrige ukene

- ▶ Syntaks og kontekstfrie grammatikker
- ▶ Maskinlæring og klassifisering
- ▶ Oblig2a: Språkmodeller og ordklassetagging (frist: **fredag** 16 oktober)
- ▶ Oblig2b: Syntaks og Semantisk klassifisering (frist: tirsdag 3 november)

I dag

- ▶ Semantikk: en lingvistisk gjennomgang
- ▶ Neste uke: semantisk språkteknologi

► Semantikk (og pragmatikk)



- ▶ Studiet av **betydning** slik det uttrykkes gjennom språk
- ▶ Betydningen til morfemer, ord, fraser og setninger
 - ▶ Leksikal semantikk
 - ▶ Setningssemantikk
 - ▶ (Pragmatikk: hvordan konteksten påvirker betydning)

Kunnskap om **betydning** (mening), men hva betyr det?

Meningsfull:

- ▶ Ord – *slem* vs *slåm*
- ▶ Setning – *Jon svømmer* vs *Jon metaforisk hver*

Flere betydninger

- ▶ Ord – *fisker*
- ▶ Setning – *Jon så mannen med kikkerten*

Samme betydning

- ▶ Ord – *sofa* vs *couch*
- ▶ Setning – *Jon ga Hanna en gave* vs *Jon ga en gave til Hanna*

Kunnskap om **betydning** (mening), men hva innebærer det?

Sannhetsbetingelser

- ▶ *All kings are male*
- ▶ *Molybdenum conducts electricity* (må kanskje sjekkes i en bok?)

Entailment (A medfører B)

- ▶ *Alfred myrdet bankdirektøren*
- ▶ *Bankdirektøren er død*

Leksikal semantikk (Ordsemantikk)

- ▶ Hvordan kan vi representere betydningen til ord?
- ▶ Hvordan er ords betydninger relatert?
- ▶ Fundamental forskjell mellom setningssemantikk og ordsemantikk (leksikal semantikk)
 - ▶ Betydninger i det mentale leksikon er **konvensjoner**
 - ▶ Setningers betydning er **komposisjonelle**
- ▶ Leksemnivå – semantiske ord
 - ▶ GITAR: *gitar, gitaren, gitarer, gitarene*
 - ▶ SYNGE: *synge, synger, sang, synget*

- ▶ Språkfilosofi: Gottlob Frege (1848–1925)
- ▶ Betydningen til et ord er dets **referanse**, koblingen til objektet det refererer til (“referenten”)
- ▶ Understreker at betydning knyttes til verden
- ▶ Men trenger noe mer:
 - ▶ Ikke alle ord referer til et objekt
 - ▶ *No baby swims*
 - ▶ Vi kan mange ord som ikke har noen referent, f.eks. *enhjørning*, *orker*, *Yoda*
 - ▶ Ord kan ha samme referanse uten å bety det samme, f.eks. *Donald Trump*, *the President*

- ▶ **Sense**: mer varig betydningsdimensjon som er separat fra referanse, kanskje en mental representasjon
- ▶ Hvordan kan vi representere sense?
 - ▶ Bilder
 - ▶ (Abstrakte) konsepter, feks semantiske trekk

- ▶ Egenskaper eller deler av ords betydning som uttrykker vår kunnskap om hva ordet betyr

Substantiver

- ▶ FEMALE: *tigress, hen, aunt, debutante, maiden, widow*, etc.
- ▶ HUMAN: *doctor, bachelor, professor, baby, teenager*

Verb

- ▶ CAUSE: *kill, uglify, darken*, etc.
- ▶ GO: *swim, crawl, throw, give*
- ▶ BECOME: *break*

- ▶ En annen måte å beskrive et ords betydning er å beskrive hvordan det forholder seg til andre ords betydning
 - ▶ **Leksikale relasjoner:** Hvordan betydninger relaterer til hverandre

Eksempel: couch

- ▶ et annet ord har samme betydning: *sofa – couch*
- ▶ et ord er mer generelt enn et annet: (*couch – furniture*)

Homonymi

Urelaterte betydninger av samme fonologiske ord (homograf/homofon)

1. Samme ordklasse

- 1.1 samme ortografi, f.eks. *lap* 'circuit of course' og *lap* 'part of body when sitting down'
- 1.2 ulik ortografi, f.eks. verbene *wring* og *ring*

2. Ulik ordklasse

- 2.1 samme ortografi, f.eks. verb *keep* og subst. *keep* (hytte)
- 2.2 ulik ortografi, f.eks. *not*, *knot*

Polysemi: flere betydninger, men betydningene er relatert (konseptuelt eller historisk)

hook noun. **1.** a piece of material, usually metal, curved or bent and used to suspend, catch, hold or pull something. **2.** short for fish-hook. **3.** a trap or snare **4.** a sharply curved spit of land, **5.** *Boxing* a short swinging blow delivered from the side with the elbow bent. ...etc

Synonymi:

Ulike ortografiske ord, men med samme, eller meget lik, betydning

- ▶ Eksempelvis *couch/sofa*, *lawyer/attorney*, *toilet/lavatory*, *large/big*
- ▶ Sjeldent fenomen i streng forstand, små nyanser i betydning
- ▶ Ulik distribusjon:
 - ▶ ulike dialekter/varianter/kontekst
 - ▶ ulikt **register**, f.eks. slang, formell, literært, etc.
 - ▶ mer eller mindre negativ: *naiv*, *godtroende*

Antonymi:

Motsatt betydning

- ▶ **Enkle** antonymer (komplementære par): positiv av det ene medfører negativ for det andre
 - ▶ *dead/alive*
 - ▶ *pass/fail (a test)*
 - ▶ *hit/miss (a target)*
- ▶ **Gradérbare** antonymer: positiv av ene medfører ikke nødvendigvis negativ av andre
 - ▶ *rich/poor*
 - ▶ *fast/slow*
 - ▶ *old/young*

Antonymi (forts.)

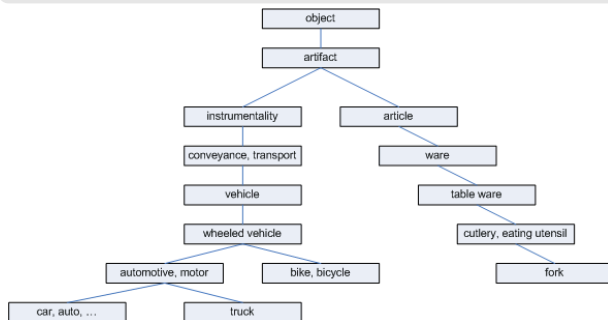
- ▶ **Reverser:** relasjon mellom ord som betegner bevegelse i motsatt retning
 - ▶ *push/pull*
 - ▶ *come/go*
 - ▶ *ascend/descend*
- ▶ **Konverser:** relasjon mellom ord som beskriver samme situasjon fra forskjellig synspunkt
 - ▶ *above/below*
 - ▶ *employer/employee*
- ▶ **Taksonomiske søstre** ord som er på samme nivå i en taksonomi (klassifikasjonssystem)

rød	oransje	gul	grønn	blå	lilla	brun
-----	---------	-----	-------	-----	-------	------

Hyponymi:

Inkluderingsrelasjon: et hyponym inkluderer betydningen til et mer generelt ord

- ▶ *hund, katt* er hyponymer av *dyr*
- ▶ *søster, mor* er hyponymer av *kvinne*
- ▶ Det mer generelle ordet kalles **hypernym**
- ▶ Basis for semantiske nettverk (f.eks. WordNet)



Meronymi:

Relasjon mellom del og helhet

- ▶ X er del av Y, Y har X
 - ▶ et *omslag* er del av en *bok*
 - ▶ en *bok* har et *omslag*
- ▶ Forskjellig fra hyponymi ved **transitivitet**:
 - ▶ hyponymi-relasjonen er alltid transitiv (*siameser* – *katt* – *dyr*)
 - ▶ meronymi-relasjonen kan være transitiv (*negl* – *finger* – *hånd*)
 - ▶ men trenger ikke være det (*arm* – *musiker* – *orkester*)

Oppsummering: <https://www.youtube.com/watch?v=awFXqSI0jR0>

Setningssemantikk

- ▶ Richard Montague (1974): **logikk** for å representere setningsbetydning
 - ▶ *Your car has been stolen* (p)
 - ▶ *Your car has not been stolen* ($\neg p$)
 - ▶ *I'll see you today or tomorrow* ($p \vee q$)
- ▶ En setnings betydning = dens **sannhetsbetingelser**
 - ▶ sannhetsverdi: hvorvidt en setning er sann eller ikke (1 eller 0)
 - ▶ sannhetsbetingelser: hvilke betingelser som må til i den virkelige verden for å gjøre en setning sann

- ▶ Semantisk relasjon mellom setninger, uavhengig av empiriske fakta
 - ▶ *The anarchist assassinated the emperor*
 - ▶ *The emperor died*
- ▶ Definisjon:
 - ▶ En setning p medfører ("entails") en annen setning q dersom det er slik at når p er sann er q sann og når q er usann så er p usann

- ▶ Entailment-relasjon er gitt ved lingvistisk struktur, trenger ikke sjekke fakta i verden for å avgjøre
- ▶ Kilden kan være leksikalsk (*assassinate* – *die*) eller syntaktisk (aktiv – passiv)
 - ▶ *Egypterne bygget pyramidene*
 - ▶ *Pyramidene ble bygget av egypterne*
- ▶ **Parafrase:** to setninger som gjensidig medfører hverandre (over)
- ▶ **Kontradiksjon:** to setninger kontradikterer hverandre dersom den ene medfører negasjonen av den andre
 - ▶ *Bankdirektøren er død*
 - ▶ *Bankdirektøren er levende*

- ▶ Semantisk analyse ved oversettelse fra engelsk (eller norsk eller ...) til et universelt metaspråk
- ▶ Førsteordenslogikk er et slikt metaspråk
- ▶ Logiske konnektiver:
 - ▶ Emil er en katt som ikke maler: $Ke \wedge \neg Me$
 - ▶ Dersom Bill drikker blir Jenny sint: $Db \rightarrow Sj$
- ▶ Kvantorer
 - ▶ Alle studenter leser: $(\forall x : Sx)Lx$
 - ▶ Alle studenter kjenner professoren: $(\forall x : Sx)Kxp$
 - ▶ Professoren kjenner (minst) en student: $(\exists x : Sx)Kpx$

Semantiske roller

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

Rolletyper

- ▶ AGENT: den som setter i gang en handling, i stand til å handle med viten og vilje (“volition”)
 - ▶ **David** *stekte kjøttet*
 - ▶ **Reven** *hoppet ut av grøfta*
- ▶ PATIENT: entiteten som påvirkes av en handling, gjennomgår ofte en forandring
 - ▶ *Edna kuttet* **buskene**
 - ▶ *Sola smeltet* **isen**

Rolletyper

- ▶ **THEME**: entiteten som blir beveget av en handling eller hvis beliggenhet beskrives
 - ▶ *David sparket **ballen** over til de andre*
 - ▶ ***Boka** er på biblioteket*
- ▶ **EXPERIENCER**: er bevisst på handlingen eller tilstanden, men er ikke i kontroll over den
 - ▶ ***Edna** følte seg dårlig*
 - ▶ ***Hanna** så røyken*
 - ▶ ***Fia** hørte ringeklokken*

Rolletyper

- ▶ BENEFICIARY: entiteten som en handling utføres for
 - ▶ *David fylte inn skjemaet for **bestemoren sin***
 - ▶ *De bakte en kake til **meg***
- ▶ INSTRUMENT: middelet som gjør at en handling kan utføres eller finner sted
 - ▶ *Hun renset såret med **et antibakterielt middel***
 - ▶ *De signerte avtalen med **den samme pennen***

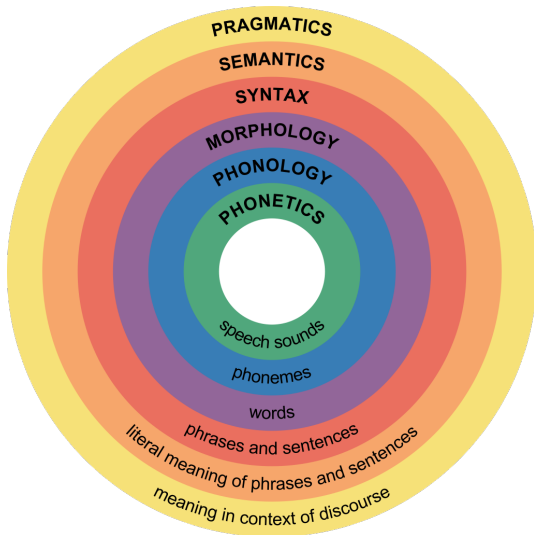
Rolletyper

- ▶ GOAL: entiteten som noe beveger seg mot (bokst./fig.)
 - ▶ *Hun leverte sertifikatet sitt til **politimannen***
 - ▶ *Hun fortalte vitsen til **vennene sine***
- ▶ SOURCE: entiteten som noe beveger seg fra (bokst./fig.)
 - ▶ *Flyet kom tilbake fra **Kinshasa***
 - ▶ *Vi fikk ideen fra et **interiørblad***

- ▶ Eksempelet: *Nina* hevet *bilen* med *jekken*
AGENT THEME INSTRUMENT
- ▶ Ikke alltid enkelt å tildele semantiske roller
 - ▶ *Fergus carried the bag to **the lighthouse*** (GOAL)
 - ▶ *Sylvie bought **him** a sports car* (BENEFICIARY)
 - ▶ ***Margareta** received a gift of flowers* (GOAL/BENEFICIARY? begge deler?)

- ▶ Del ut følgende semantiske roller for setningene under:
 - ▶ PATIENT: entiteten som påvirkes av en handling, gjennomgår ofte en forandring
 - ▶ AGENT: den som setter i gang en handling, i stand til å handle med viten og vilje
 - ▶ INSTRUMENT: middelet som gjør at en handling kan utføres eller finner sted
- 1. *Jon knuste vinduet*
- 2. *Jon knuste vinduet med en stein*
- 3. *Steinen knuste vinduet*
- 4. *Vinduet ble knust av Jon*

- ▶ Semantiske roller identiske i parafraser
 - ▶ *The dog bit the stick/the stick was bitten by the dog*
- ▶ Kan også være identiske i setninger som ikke er parafraser
 - ▶ *The boy opened the door with the key*
 - ▶ *The key opened the door*
 - ▶ *The door opened*



- ▶ Studiet av vår språklige forståelse i **kontekst**
 - ▶ lingvistisk kontekst, feks et dokument
 - ▶ pronomener
 - ▶ ikke-lingvistisk kontekst
 - ▶ *Kan du sende saltet?*

Pronomen:

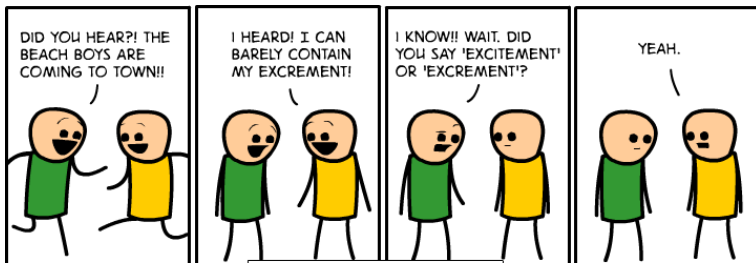
jeg, hun, dere, seg, hverandre, hvem, man

1. av svært ulik form, uregelmessig bøyning
2. som substantiv, kan fungere som setningsledd alene
3. lite eget innhold, får betydning fra sammenhengen (*konteksten*)
 - ▶ *Jeg* liker grammatikk
 - ▶ *Man* skal respektere *hverandre*
 - ▶ *Hvem* tok vesken?

- ▶ Pronomen får sitt innhold enten fra et element i selve talesituasjonen eller fra en nominal leddtype (typisk substantiv) i konteksten.
- ▶ Leddet som gir pronomen innhold er pronomenets **antesedent**
 - ▶ Se *her*
 - ▶ Er *du* sulten?
 - ▶ Gro Harlem Brundtland er en tidligere norsk politiker. *Hun* var norges første kvinnelige statsminister. . .

- ▶ **Deiksis**: ord som typisk får referansen fra den ikke-lingvistiske konteksten
 - ▶ *this person*
 - ▶ *that man*
 - ▶ *these women*
- ▶ Tids-og stedsdeiksis
 - ▶ *nå, i morgen, forrige uke*
 - ▶ *her, der, dette stedet, denne parken*

- ▶ Grice: kjent språkfilosof
- ▶ Samtaler er basert på prinsipp om **samarbeid**
- ▶ **Maksimene** til Grice: uskrevne regler som muliggjør en sammenhengende og meningsfull samtale
 - ▶ **kvantitet**: si hverken mer eller mindre enn nødvendig
 - ▶ **relevans**: vær relevant
 - ▶ **måte**: vær klar og konsis
 - ▶ **kvalitet**: ikke lyv



Cyanide and Happiness © Explosm.net

Kvantitetsmaksimet:

- ▶ *Hva skjedde i dag?*
- ▶ *Møtet startet tre minutter for sent, jeg satt ved døren, den første som sa noe var ...*
- ▶ *Ikke noe*

Relevansmaksimet:

- ▶ *Hvordan har du hatt det i dag?*
- ▶ *Jeg hater tomater*

Måtesmaksimet:

- ▶ *Liker du skjorten min?*
- ▶ *Den er interessant*

Kvalitetsmaksimet:

- ▶ Venn spør: *Vet du hvor bilnøkklene mine er?*, og forventer et sannferdig svar

- ▶ Taleakter: språk kan brukes til å utføre handlinger, f.eks. love noe, vedde, advare, døpe noen, gratulere noen, sverge, etc.
 - ▶ *Jeg vedder 100 kroner på at Amman vinner*
 - ▶ *Møtet er hevet*
 - ▶ *Jeg erklærer dere herved for rette ektefolk*
 - ▶ *Jeg lover å komme tilbake*
- ▶ Underliggende målet med ytringen varierer med konteksten
 - ▶ *Bursdagsfest 11/4 kl 20*

- ▶ Studiet av betydning slik det uttrykkes gjennom språk
- ▶ **Ord:** leksikal semantikk
 - ▶ referanse vs sense
 - ▶ leksikale relasjoner (*nymi)
- ▶ **Setning:** formell semantikk
 - ▶ sannhetsverdi
 - ▶ logikk
- ▶ **Setning:** semantiske roller
 - ▶ hvilke roller de forskjellige deltagerene inntar
 - ▶ abstrakt beskrivelse av situasjonen verbet angir

- ▶ Pragmatikk: språklig forståelse i **kontekst**
 - ▶ kontekst gir referanse/innhold
 - ▶ pronomen
 - ▶ deiksis
 - ▶ sosial kontekst med samarbeidsprinsipper bidrar til forståelse (Grice)
 - ▶ ikke-lingvistisk kontekst påvirker betydning (taleakter)