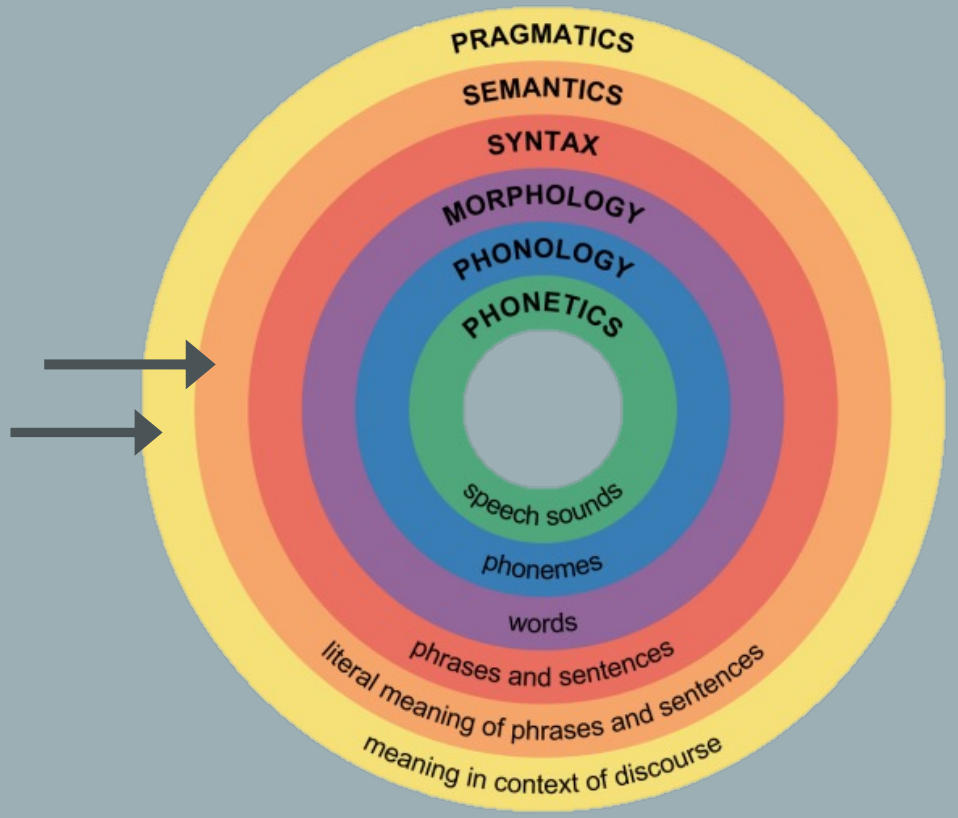
SEMANTIKK (OG LITT PRAGMATIKK)



HVA ER SEMANTIKK?

Studiet av betydning slik det utrykkes gjennom språk

Kunnskap om betyding- hva betyr ord og setninger?

- Ser ikke på hvordan konteksten bidrar til meningsinnholdet, slik som i pragmatikk

Leksikal semantikk: Betydningen av ord

Setningssemantikk: Betydningen av setninger

ORD OG SETNINGER KAN:

Ha flere betydninger:

- *Pære* 🍐 💡?
- *Jeg slo Ola med Boka*

Ha de samme betydningene:

- *Sofa* vs. *couch*
- *Jeg drepte noen* vs. *Noen ble drept av meg*

SETNINGER KAN:

Ha sannhetsbetingelser:

- *Alle barn er yngre enn voksne*

Medføre at en annen setning er sann eller usann («Entailment»):

- A: *Dronningen drepte kongen*
  - B: *Kongen er død*
  - A medfører B

LEKSIKALE RELASJONER

Beskriver et ords betydning ved å beskrive hvordan det forholder seg til andre ords betydning.

Homonymi

Ord som har samme uttale eller skrivemåte, men ulik betydning

Merke (verb) vs. Merke (subst.)

Samme ortografi, ulik ordklasse

Bli vs. Blid

Ulik ortografi og ordklasse

Pære 💡 *vs. Pære* 🍐

Samme ortografi og ordklasse

Polysemi

Flere betydninger men betydningene er relaterte.

Stjerne 🌟 *vs. Stjerne* 🦹

Kan vise til et himmelegeme eller en berømt person. Beslektede betydninger.

Meronymi

Relasjon mellom del og helhet

En *finger*☝ er en del av en *hånd*🖐

Antonymi

Motsatt betydning

<u>Enkle antonymer</u>

Død/levende

<u>Gradérbare antonymer</u>

Gammel/ung

Hyponymi

Et hyponym inkluderer betydningen til et mer generelt ord

Hund, katt er hyponymer av hypernymet *dyr*.

Synonymi

Ulike ortografiske ord, men med samme eller veldig lik betydning

Bra vs. God

SEMANTISKE ROLLER

Aspekt ved setningsbetydning: Hvilke roller de forskjellige deltagerne har

Ole følte seg dårlig. Elise trøstet ham med mat.

Experiencer
Er bevisst på handlingen eller tilstanden, men har ikke kontroll over den

Pasient
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

FLERE SEMANTISKE ROLLER

Theme

Entiteten som blir beveget av en handling eller hvis beliggenhet beskrives

IFI ligger ved Forskningsparken

Goal

Entiteten som noe beveger seg mot

Hun leverte eksamen til sensoren

Beneficiary

Entitet som en handling utføres for

Jeg handlet for bestemoren min

Source

Entiteten som noe beveger seg fra

Jeg fikk svaret fra en venn

SEMANTIKK I SPRÅKTEKNOLOGI

Hva slags oppgaver kan løses innen semantikk med språkteknologi?

Tre språkteknologiske felt/oppgaver for hvert nivå innen betydning:

1. Ord: Word Sense Disambiguation (WSD)
2. Fraser: Named Entity Recognition (NER)
3. Setninger: Semantic Role Labeling (SRL)

WSD (BETYDNINGSDISAMBIGUERING)

Disambiguering: Analyse av flertydighet

Gitt en setning med et spesifikt målord og en liste med betydninger, angi korrekt betydning for målordet i setningen

«Jeg dro til en bank i dag»

✅ Bank: substantiv, forretningsinstitusjon

❌ Bank: substantiv, banking, pryl

❌ Bank: verb, å banke (på)

Klassifisering basert på et annotert datasett

NAMED ENTITY RECOGNITION (NER)

Automatisk gjenkjenning og kategorisering av egennavn

PERSON GPE ORG

Jonas Gahr Støre er statsminister i Norge, og sitter på Stortinget

NE Type	Eksempler
ORGANIZATION	Omnicom, WHO
PERSON	George Washington, President Obama
LOCATION	Downing St., Mississippi River, Norway
DATE	June, 2011-05-03, 03/05/2011
TIME	two fifty a.m., 1:30 p.m.
MONEY	175 million Canadian Dollars, GBP 10.40
FACILITY	Washington Monument, Stonehenge
GPE	Washington D.C., Norway

SEMANTIC ROLE LABELING (SRL)

Gitt et predikat i en setning, finn dets semantiske roller

[*arg0* Læreren] leste [*arg1* en bok]

[*arg1* Boka] ble lest av [*arg0* læreren]

Læreren er alltid *agent* og boka er alltid *instrument*

Disse oppgavene har noe til felles: Veiledet klassifisering

- Maskinen får se ferdigannotert data (treningsdata) før den klassifiserer på egenhånd

Bruker trekk for å representere treningsdataen:

-Ord

-Lemma

-ordklasse

-NE-klasse

-Syntaktisk kategori

..osv..

Ikke-veiledet klassifisering: Maskinen finner mønster uten forhåndsannotert data

NAIVE BAYES-KLASSIFISERING

Populær metode for klassifisering innenfor WSD (men også innen mye annet)

Bruker informasjon om ord i konteksten for disambiguering av betydning

1. Prior-sannsynlighet for betydningen P(s)

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

2. Sannsynlighet for individuelle trekk P(fj|s)

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

BIO-KLASSIFISERING

Metode for klassifisering innenfor NER. Ord-for-ord klassifisering.

B: begynnelsen av egennavn, I: Innenfor egennavn, O: Utenfor (outside) egennavn

B_PERS	I_PERS	I_PERS	O	O	O	B_GPE	O	O	O	B_ORG
Jonas	Gahr	Støre	er	statsminister	i	Norge	og	sitter	på	Stortinget

SYNTAKTISK ANALYSE

Metode for klassifisering innenfor SRL

Klassifiserer noder (konstituenter) i det syntaktiske treet

