IN1140 – Uke 10

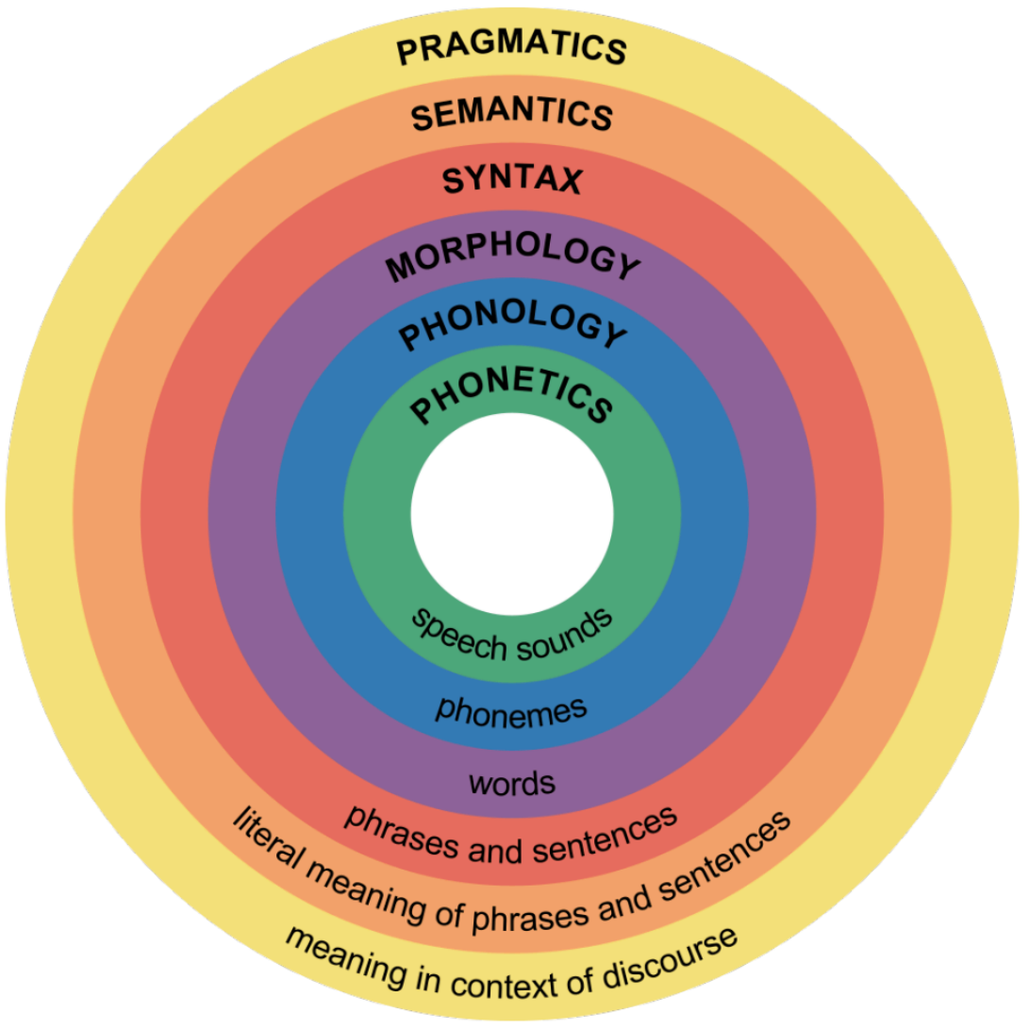
Siste oblighjelp🎉

+

Hva enn dere vil🤗

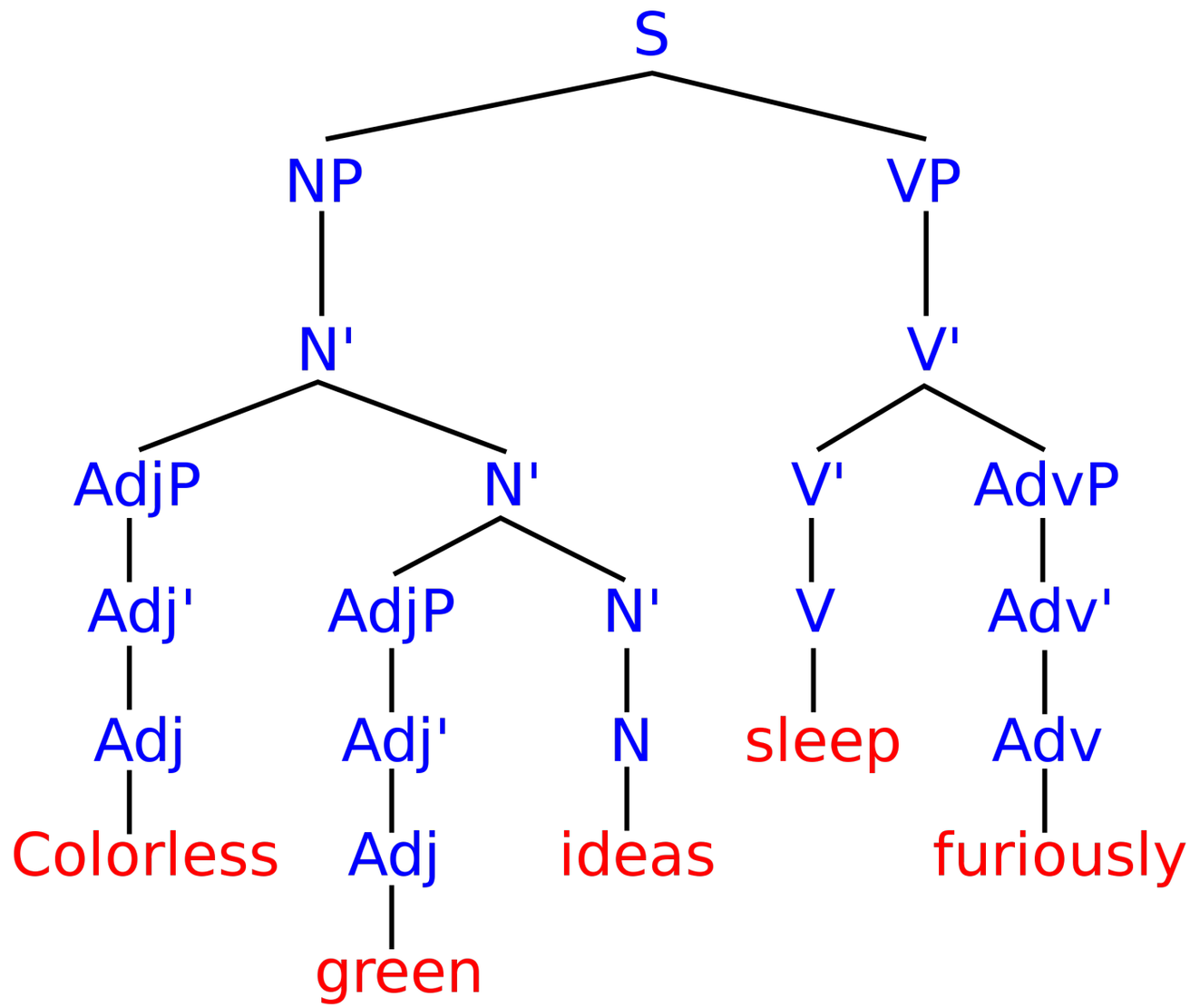
Semantikk

- «Studiet av betydning slik det uttrykkes gjennom språk»
  - Kunnskap om ord og setninger
- Vi skiller mellom:
  - Leksikalsk(/ord) semantikk
  - Setningssemantikk



Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo

Eksempel på leksikalsk flertydighet



«Fargeløse grønne ideer sover rasende»

Noam Chomsky (1957)

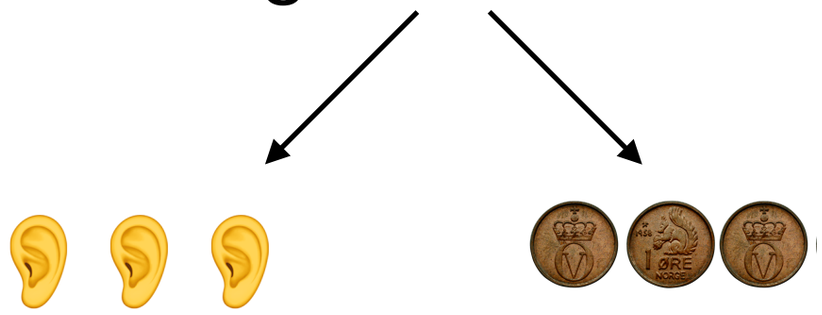
Setningen mangler semantisk innhold



Semantikk i språkteknologi

Word Sense Disambiguation (WSD)

- Disambiguering: Å avklare tvetydighet
- Vi ønsker å finne riktig betydning for et ord i en setning
  - «Det vil koste deg 3 øre»



- Her kan vi f.eks. bruke Naive Bayes —> $P(tolkning)\prod_{j=1}^{n} P(trekk_i \mid tolkning)$

Named Entity Recognition (NER)

- Automatisk gjenkjenning og kategoriskering av egennavn
- BIO-tagging:
  - B-??? – Beginning
  - I-??? – In
  - O – Out

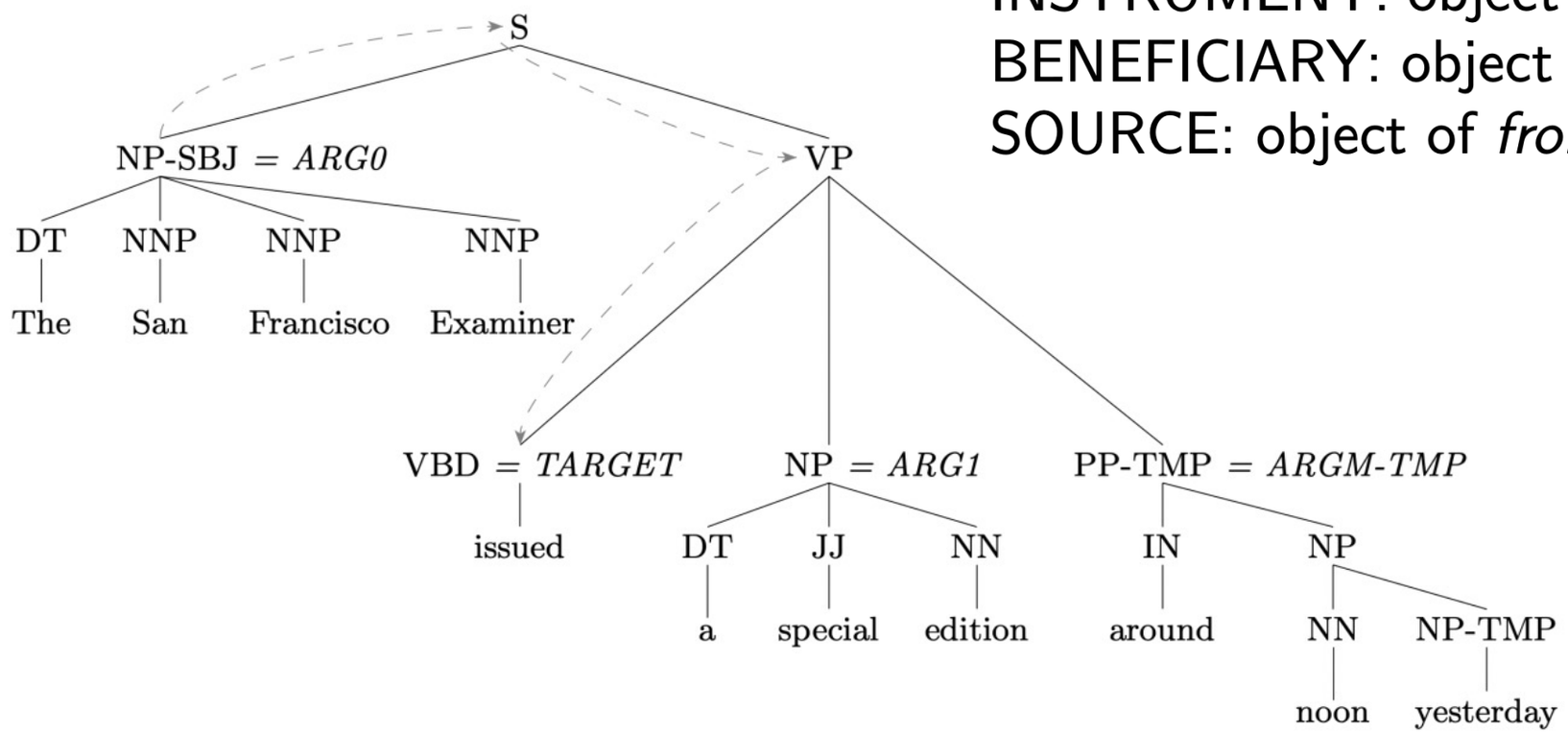
B-PERS I-PERS I-PERS O O O B-GPE O O O B-ORG

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

Semantic Role Labeling (SRL)

- Automatisk rolletagging i setninger
- Gjøres ofte via syntaktisk analyse

AGENT: subject
PATIENT: direct object
INSTRUMENT: object of *with*
BENEFICIARY: object of *for*
SOURCE: object of *from*



Semantiske roller

For å forstå setningen, må vi forstå hva slags roller entitetene har

- **Agent** – Den/det som handler med viten og vilje
- **Patient** – Den som påvirkes av handlingen
- **Theme** – Det som plasseres/beveges av handlingen
- **Experiencer** – En som er bevisst på, men ikke i kontroll av handlingen
- **Beneficiary** – Den som handlingen utføres for
- **Instrument** – Det middelet for å utføre handlingen
- **Goal** – Det som noe beveger seg mot
- **Source** – Den som noe beveger seg fra

Oppgave! Gi roller:

- **Agent** – Den/det som handler med viten og vilje
- **Patient** – Den som påvirkes av handlingen
- **Theme** – Det som plasseres/beveges av handlingen
- **Experiencer** – En som er bevisst på, men ikke i kontroll av handlingen
- **Beneficiary** – Den som handlingen utføres for
- **Instrument** – Det middelet for å utføre handlingen
- **Goal** – Det som noe beveger seg mot
- **Source** – Den som noe beveger seg fra

Jeg fikk et eple.

Beneficiary

Jeg kommer fra **Oslo**.

Source

Han knuste døra med **en øks**.

Instrument

Leksikalske relasjoner

Synonymi

- Ulik ortografi (stavemåte), men (ca.) samme betydning
  - «TV» = «fjernsyn»
  - «katt» = «pus»
  - «kunne» ≈ «mestre»

Antonymi

- Motsatt betydning
  - Enkle antonymer: «død» ≠ «levende», «på» ≠ «av»
  - Graderbare antonymer: «ny» <—> «gammel»

- To "synonyme" setninger kaller vi **ekvivalente**
  - «Han ga henne en klem» ≡ «Hun fikk en klem av ham»

Leksikalske relasjoner

- Homonymi – Samme uttale eller ortografi, men ulike ord
  - Homofon – Samme uttale, f.eks. «jul» og «hjul»
  - Homograf – Samme ortografi, f.eks. «får» (V) og «får» (S)
- Polysemi – Relaterte betydninger av homografer
  - «En vill plante» og «En vill konsert»

Leksikalske relasjoner

- Meronymi – Relasjon mellom del og helhet
  - «skjegg» < «julenisse», «side» < «bok», «negl» < «tå» < «fot»
- Hypernymi – Generalisering av ord (motsatt vei -> homonymi)
  - «menneske», «hund», «slange» ∈ «dyr» (hypernymet)

Leksikalske relasjoner

Sannhet fra utsagn (hvis A, så B)

A: Jeg svømmer

B: Jeg er i en væske

PRAGMATIKK

*Studiet av vår språklige forståelse i **kontekst***

PRONOMEN

«Jeg», «han», «de», «den», «det», +++

Typiske eksempler på ord som først får betydning i en gitt kontekst

GRICE SINE MAKISMER:

I en samtale forventer vi:

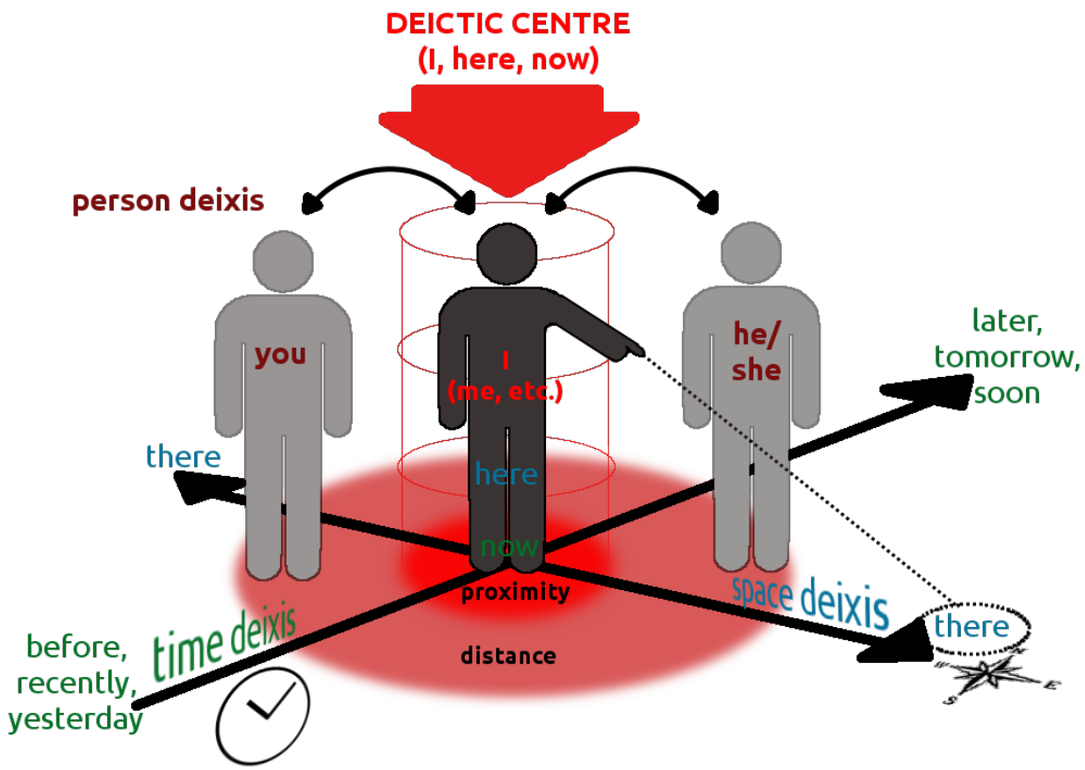
Kvalitet: Sannferdighet

Kvantitet: Ikke for mye eller for lite

Relevans: Relevant for samtalen

Måte: Kort, konsis, tydelig

DEIKSIS



PRAGMATIKK

*Studiet av vår språklige forståelse i **kontekst***

GRICE SINE MAKISMER:

Grice mente at man i en rasjonell, samarbeidende samtale må forholde seg til fire normer for at samtalen skal fungere til sitt formål (god kommunikasjon). Den som taler må etterstrebe å følge disse normene, mens den som lytter må anta at taleren følger dem.

De fire normene er:

Kvalitet – Ikke lyv. Dette inkluderer påstander du ikke har dekning for.

Kvantitet – Ikke si for lite til å bli forstått, men ikke forvirr med unødvendig informasjon.

Relevans – Det du sier bør være relevant for samtalen.

Måte – Ha tydelighet i språket. Unngå vaghet og tvetydighet.

Neste uke

- Eksamensforberedelser
- Samle trådene
- Gamle eksamensoppgaver
- Deres spørsmål