

## Morfem

Dette er grunnstammen i et ord og som sammen utgjør ordene. F.eks så er adjektivet penere en sammensetning av morfemene pen og ere. Ved å legge til en spesifikk morfem til et annet endrer man ordklassen ordet tilhører. F.eks «ete» gjør et verb til et adjektiv.

- Oppdeling:
  - Affiks
    - Prefiks
      - En morfem som kommer før ordstammen. «un» i «unhappy» er et prefiks til roten «happy»
    - Post/suffiks
      - Morfem som kommer etter roten. Eks «ere» i «penere»
    - Infiks
      - Morfem som er lagt til inni roten
    - Surkumfiks
      - Morfem som legges til både i slutten og i starten av roten
    - Interfiks
      - Morfem som kommer mellom to røtter f.eks «e» i «hønsegården»
  - Bundne og frie
    - Fri morfemer
      - Er morfemer som kan stå alene. De har en betydning i seg selv. Eks «pen» er et slikt morfem.
    - Bundne morfemer
      - Er morfemer som ikke kan stå alene, de må komme sammen med andre morfemer som en eller slags affiks til det frie morfemet (roten). Et eksempel her er adjektivendelsen «ere» som i penere.
- Stamme og rot
  - Stamme er den edelen av ordet som står uendret. Den trenger ikke være et eget morfem men kan være en sammensetning av flere
  - Roten er en del eller hele stammen i enkelte tilfeller. Roten er kun et morfoem, det som utgjør grunnbolken av mening.
    - Eks:
      - Ta ordet alvorligere
        - alvorlig er stammen
        - alvor er roten
- Avledningsmorfem og bøyingsmorfem
  - Avledningsmorfem:
    - gjerne et bundet morfem.
    - Ved å legge et avledningsmorfem til en stamme vil det bli konstruert et nytt ord.
    - Eks stammen «brun» + avledningsmorfemet «het» = «brunhet»
  - Bøyingsmorfem:
    - Minner veldig mye om avledningsmorfemene men istedenfor å lage et helt nytt ord så lager den en ny variant av det samme
      - Tilfører ordet mer informasjon som f.eks hvor mye eller hvor mange
      - Eks «ere» i «penere» er en bøyingsmorfem

## Syntax

Viten rundt setningsoppbygging er syntaks. f.eks det at på Norsk står verbet alltid i andre posisjon.

- Struktur
  - Måten på hvordan en setning er bygget opp rent strukturelt gir forskjellige betydninger. F.eks to like setninger men med forskjellig setningsoppbygging vil ha forskjellig syntaks og derfor også forskjellig betydning. I tilfeller som dette kalles setningen for tvetydig eller ambiguous
  - Det som viser til strukturen til en setning er parsetreet den er laget av som kommer fra CFGen
  - Grammatikk
    - En grammatisk setning er en setning vi kan tildele en struktur avhengig om setningen gir mening eller ikke rent språklig.

- Konstituent
  - En del av en setning som fungerer som en egen enhet.
  - Tester for å finne ut om noe er en konstituent:
    - Kan den stå alene?
      - Eks «hunden spiste leksene mine» spør «hva spiste hunden» svar «leksene mine» altså er «leksene mine» en konstituent
    - Kan deler av setningen erstattes med pronomener?
      - «Hva spiste han?» Han = hunden, altså er «hunden» en konstituent
    - Kan enheten flyttes til helt et annet sted i setningen?
      - «Leksene mine ble spist av hunden»
- CFG
  - Er reglene som utgjør syntaksen
  - Det bak parentesene er regler som ikke gjelder alle verb hvorav det før er annsett som genereller regler.  
 S -> NP VP  
 NP -> Det N (| NP PP)  
 VP -> V NP (| V | V NP PP | V NP NP | VP PP)  
 N -> «walk» | «talk» | ++  
 Det -> «the» | ++

## Semantikk

Hvordan vi uttrykker mening gjennom det vi sier. Det er to hovedtyper leksikalsksemantikk og setningssemantikk, disse er veldig forskjellige

### Leksikalsksemantikk

- Betydningen av ord
- Referanse
  - Et ord i seg selv har ingen betydning. Ordet får først sin betydning ved at vi forbinder det med noe annet. Dette kalles referansen til ordet. Eks statsminister sier ingenting før vi forklart at det er overhode i den norske stat. Altså før vi får noe å referere statsminister til
  - To ord med samme referanse trenger ikke være like, ta f.eks president og Obama
- Semantiske trekk
  - En måte å gruppere ord på
  - Ord blir gruppert basert på det de har tilfelles
- Homonymi
  - En annen måte å dele inn ord på
  - Deles inn etter hvordan de staves eller høres ut, trenger ikke ha noe med hverandre å gjøre
    - Eks hund og hun
- Polysemi
  - Kan minne om homonymi en er ikke det samme
  - Et ord som har flere betydninger hvor betydningen gjerne stammer fra hverandre, eller i det minste er relatert.
    - Eks face, kan refere til det man har i ansikte eller f.eks det som er på en klokke
- Synonymi
  - Inndeling hvor en gruppe består av ord som har samme betydning
- Antonymi
  - To ord har motsatt betydning av hverandre
    - Enkle (komplementære par)
      - Har du det ene så har du ikke det andre
        - Liv/Død
    - Graderbare
      - Det kan settes på en skala og at fravær av en ting vil ikke nødvendigvis implisere den andre
        - Rik/Fattig
    - Reverser
      - Bevegelse i forskjellige retninger

- Dytte/Dra
- Konverser
  - Synspunkt fra forskjellig sider
  - Få/Gi
- Taksonomiske søstre
  - Kan settes inn i et taksionom, klassifiserings system
  - Der forskjellige ord er på samme nivå
  - Eksepel her er farger
- Hyponymi
  - Der et ord har en mer spesifikk betydning enn et annet ord (hypernymet) eks katt er et hyponymi til hyponymet dyr
  - Kan tenke på dette som klassehierarki i Java
    - Subklasser extender superklasser
- Meronymi
  - Som hypomymi men hvor mengden er transitiv
    - $x \sim y \wedge y \sim z \Rightarrow x \sim z$

### Setningssemantikk

- Betydningen av en setning
  - Sannhetsbetingelse
    - Basert på leseren referanser kan setningen gjøres sann? Isåfall hva er betingelsene?
  - Førsteordens logikk brukes her
- Predikat
  - De ordene eller tingene som kan gjøre setningen sann.
- Problemer
  - Metaforer, et ord skal ikke tolkes i sin faktiske betydning men i en som kulturen har blitt enig i over tid
  - Faste uttrykk i språk
    - Slike setninger vil alene være meningsløse men sett i større sammenheng gi mening
      - På denne måten får vi en gråsoner mellom ord og setninger. Hva er ordets betydning og hva er ordets betydning sett i lys av en setting?
        - Ord er konvensjonelle i sin betydning
        - Setninger er komposisjonelle
- Semantiske roller
  - Agents
    - Den som utfører handlingen, med vilje
  - Patiens
    - Den som påvirkes av handlingen, blir ofte endret
  - Theme
    - Den som blir beveget eller beskrevet
  - Experierer
    - Den som opplever handlingen, men gjerne uten vilje
  - Benerficiary
    - Den som drar nytte av handlingen
  - Intrument
    - Det handlingen utføres med
  - Goal
    - Målet bak handlingen, beveger seg «til»
  - Source
    - Der handlingen stammer fra
- Pragmatikk
  - språk i kontekst
    - Både linguistisk og ikke-lingvistisk
  - Anafor og deiksis
    - Anafor: Peger tilbake til noe som ble oppgitt tidligere. F.eks pronomen
    - Deiksis: Peger på noe ikke-lingvistisk

- Kan referere til ting men også til tid/rom
- Grice
  - Kjent filosof som så på språklige konvensjoner. Kom opp med følgende 4
    - Ikke si mer eller mindre enn nødvendig
    - Ikke lyve
    - Relevant
    - Skal sies på en klar og forståelig måte
- Speech acts
  - Når en setning utfører en handling fremfor å være en abstrakt strøm av ord
    - Eks møtet er hevet
  - Eller den kan implisere at en handling må gjøres
    - Jeg er sulten informerer ikke om at du er sulten men er en oppfordring på å skaffe mat

#### Sammenheng mellom ord

- Parafrasering
  - omskrivingen av en setning slik at den holder si betydning men bruker andre ord
- Temporale uttrykk
  - når i tid forekommer en hendelse.
  - «For en uke siden, dagen før bryllupet, døde bestefar». Når døde bestefar?
- Entailment
  - Det forventede svaret som følge av et spørsmål
  - Vi sier svaret entails spørsmålet
- Metaforer
- Negasjoner
- Spekulasjon og faktivitet

#### Metoder

- Klassifisering
  - Gitt et ord og en liste over mulige betydninger, velg en betydning
  - Gitt et predikat i en setning, finn den semantiske rolle
- automatisk tilegning av informasjon
  - Hvilke ord og fraser betyr det samme
  - del-helhet relasjoner
  - automatisk gruppere setninger som inneholder et vist ord basert på betydning

#### WSD

- WordNet utgangspunkt for WSD
  - Merke forekomst av ord med riktig betydning
  - Generaliserer ord, kat >> dyr
  - Viktig innenfor språkteknologi
    - Gir setningen et målord og en liste med betydninger
      - Angir den betydningen som er best for målordet gitt setningen
    - Klassifiserer basert på anotert datasett
  - Statistisk klassifisering
    - sentral metode innenfor maskinlæring
    - automatisk avgjør hvilken kategori en observasjon tilhører
  - Steg 1
    - Hent ut trekk som er sentrale i diasambiotetn i treningsdataen
      - Eks ta ordet SKIM (eng) kan bety «lese fort over noe», «fjerne skum fra melk» eller «bevegelse». Gitt en setning skal vi finne ut hvilken av disse setningene bruker.
      - Trekk: BAG-OF-WORDS
        - Henter ut alle ord som er i setningen som det ordet vi vil finne ut av
          - ikke ordnet
          - og gir dem en sence basert på hva ordet betyr i den konteksten
          - Gjør dette med flere setninger. Hvor hver setning blir sin liste
          - kontekst til målordet kan representeres ved
            - lemma
            - ordformer

- ordklassetagger
- kombo av det over
- Trekk
  - trekkvektor:  $[w_{i-2}, POS_{i-2}, w_{i-1}, POS_{i-1}, w_{i+1}, POS_{i+1}, w_{i+2}, POS_{i+2}]$ 
    - eks setningen «Remove the vanilla pod, SKIM the jam, and let it cool: Remove
      - trekkvektor = [vanilla, JJ, pod, NN, the, Det, jam, NN]
- Vektormodell
  - Teller forekomsten av et ord i et dokument og setter det på si repsektive plass i vektoren
  -
- Trekk: N-gram
  - Tar de N orden på hver side av målordet og lager en vektor ut av dem. Er det ingen der lager vi «dummy» ord som tar plassen
- Klassifisering
  - Gitt teningsdata og trekkvektorer kan vi vha algoritmer som NAIVE BAYES trene en klassifiserer
- NAIVE BAYES
  - $\hat{s} = \operatorname{argmax} P(s|f)$ 
    - $\hat{s}$ : mest sansynlige tolkingen
    - s: en tolkning
    - f: en trekkvektor
    - $P(s|f) = P(f|s) \cdot P(s) / P(f)$
  - Omformuleres til:
 
$$\hat{s} = \operatorname{argmax} P(f|s)P(s)/P(f)$$
  - Ved at dataen er uavhengig kan vi forenkle til
 
$$\hat{s} = \operatorname{argmax} P(s) \prod_i P(f_i|s)$$
    - $P(s) = \text{count}(s, w) / \text{count}(w)$
    - $P(f|s) = \text{count}(s, f) / \text{count}(s)$

## Ordklasser

- Ordklassekriterier
  - Vi trenger enkelt kriterer for å kunne avgjøre hvilke ordklasser de forskjellige ordene hører til i
  - Morforlogisk = Formelle kriteret
    - Inndeling basert på hvordan ordet er bygget. Ta f.eks penere. Vet at ord som slutter med morfemet «ere» er adjektiv.
  - Syntaktisk = Funkjonelt kriteret
    - Inndeling basert på hvor i setningen et ord av en type ordklasse pleier å stå
    - Med hvilke ordklasser kan en ordklasse kombineres
      - På Norsk f.eks pleier verbet alltid å stå i andre posisjon i setningen.
      - Verb har veldig ofte en N forran seg
  - Semantisk = Betydningskriteret
    - Inndeling basert på betydningen av et ord. Hva er det ord som tilhører en ordklasse har til felles når det kommer til betydningen av orde?
      - Eks ordklassen verb innholder ord som beskriver handlingen i en setning. Enten at noen beveger seg eller at noe tenker