

# INF1820: Introduksjon til språk-og kommunikasjonsteknologi

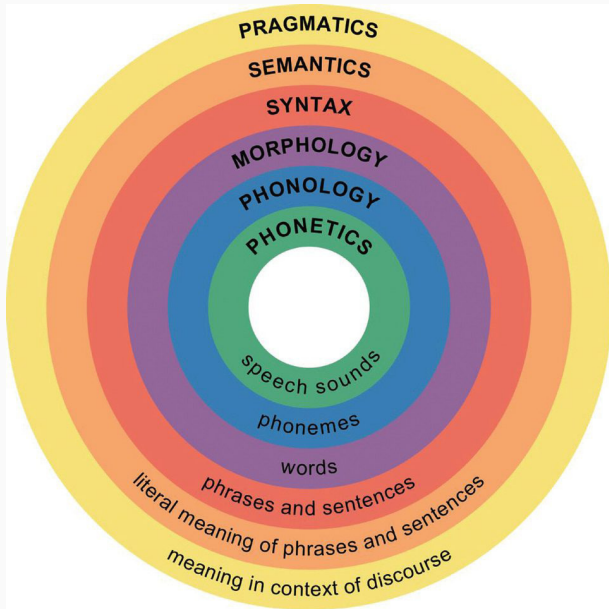
Åttende forelesning

---

Lilja Øvrelid

13 mars, 2017

# I dag: Syntaks



# Bakgrunn

- Syntaks: studiet av prinsipper og regler for setningsdannelse
- Beskrivelse av naturlige språk via sett med generelle (universelle) regler
- Gammel disiplin
  - Pāṇini: sanskrit grammatiker (4 BC)

- Syntaks: studiet av prinsipper og regler for setningsdannelse
- Beskrivelse av naturlige språk via sett med generelle (universelle) regler
- Gammel disiplin
  - Pāṇini: sanskrit grammatiker (4 BC)
- Mange teorier: mye til felles, en del forskjeller
  - Government and Binding (GB)
  - Minimalism Program (MP)
  - Head-driven phrase structure grammar (HPSG)
  - Lexical Functional Grammar (LFG)
  - Categorical Grammar
  - Dependency Grammar
  - ...

- Noam Chomsky: *Syntactic Structures* (1957)
- Syn på språk inspirert av naturvitenskap: språket er en mengde setninger
  - **mengde:** en samling objekter, f.eks.:  $\{a, b, c\}$ ,  $a \in \{a, b, c\}$ ,  $d \notin \{a, b, c\}$

- Noam Chomsky: *Syntactic Structures* (1957)
- Syn på språk inspirert av naturvitenskap: språket er en mengde setninger
  - **mengde:** en samling objekter, f.eks.:  $\{a, b, c\}$ ,  $a \in \{a, b, c\}$ ,  $d \notin \{a, b, c\}$
- Eksplisitt beskrivelse av denne mengden
  - *Alle nordmenn liker ikke fotball*
  - *Liker ikke alle nordmenn fotball*
  - *\*Nordmenn alle ikke liker fotball*
  - *\*Liker nordmenn fotball ikke alle*

- Generativ?
  - den grammatiske beskrivelsen skal **generere** alle og bare de grammatiske setningene i språket.
    - *Alle nordmenn liker ikke fotball*
    - *Liker ikke alle nordmenn fotball*
    - *\*Nordmenn alle ikke liker fotball*
    - *\*Liker nordmenn fotball ikke alle*

- Ekstensjonell (*oppregning*) vs. intensjonell (*beskrivelse*):
  - ekstensjonell: liste opp alle setninger:

*{... Alle nordmenn liker fotball,  
Alle nordmenn liker ikke fotball,  
Alle nordmenn liker fotball og friidrett,  
Alle nordmenn liker ikke fotball og friidrett... }*



- Problem med ekstensjonell beskrivelse:
  - i prinsippet uendelig antall setninger
  - sier ingenting om setningers **struktur**
    - *Alle nordmenn*
    - *liker fotball*
    - *Alle*
    - *nordmenn*
    - *liker*
    - *fotball*

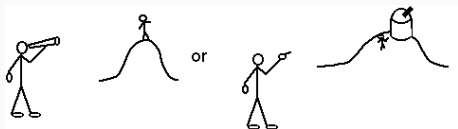
# Grammatiske beskrivelser

- En **grammatikk** er en beskrivelse av et språk
- Inneholder:
  - hvordan en språklig enhet ser ut – **form**
  - rollen enheten spiller i forhold til andre språklige enheter – **funksjon**
- Samme form kan ha ulike funksjoner:
  - *Hun kjøpte et **pent** hus.*
  - *Huset er **pent**.*
  - *Hun synger **pent**.*
- Form og funksjon for større enheter enn ord
  - **Et pent hus er ikke alltid dyrere.**
  - **Vi bodde i et pent hus.**

# Grammatiske beskrivelser

## Setningers struktur:

- Mer enn bare den lineære rekkefølgen
- Hierarkisk ordnede **fraser**:
  - funksjon - enkeltord og større grupper:
    - *Jeg liker fotball*
    - *Alle nordmenn liker fotball*
  - flertydighet - samme (lineære) ordstreng, ulik frasestruktur:
    - *Hun så gutten med teleskopet*
    - [*så gutten med teleskopet*]
    - [*gutten med teleskopet*]



- En generativ grammatikk:
  - genererer alle og bare de grammatiske setningene i et språk
  - tildeler hver setning en strukturell beskrivelse
- En (generativ) grammatisk beskrivelse er en eksplisitt redegjørelse for mengden setninger i et språk
- Syntaktiske regler

# Generativ grammatikks metodologi

- Data: våre språklige intuisjoner
- Hypotetisk-deduktiv metode
  - hypotetisk: formulerer hypoteser
  - deduktiv: deduserer (utleder) testbare konsekvenser fra hypotesene
  - hypoteser kan ikke verifiseres, men styrkes

# Generativ grammatikks metodologi

- Data: våre språklige intuisjoner
- Hypotetisk-deduktiv metode
  - hypotetisk: formulerer hypoteser
  - deduktiv: deduserer (utleder) testbare konsekvenser fra hypotesene
  - hypoteser kan ikke verifiseres, men styrkes

- Den svenske professoren (en vits):

*Det var en gang en svensk professor som arbeidet ut i fra en hypotese om at lopper hører med beina. Han hadde en dressert loppe. Når professoren sa "Hopp!", så hoppet loppen. Så gjorde professoren følgende forutsigelse: "Hvis jeg river beina av loppen, blir den døv." Han gjorde det. Når han deretter sa "Hopp!" til loppen, hoppet den ikke. Professoren konkluderte som følger: "Hypotesen om at lopper hører med beina stemmer."*

# Syntaks

---

“Studiet av hvordan setninger bygges opp av ord og ordkombinasjoner, såkalte konstituenten”

- **Syntaktisk form** - konstituenten beskrives i form av ordklasser, fraser:
  - fraser - større konstituenten over ordnivå
  - fraser navngis etter **hodet** - det sentrale, obligatoriske medlemmet, referanse
- **Syntaktisk funksjon** - konstituenten beskrives i form av sin funksjon i setningen som helhet
  - Subjekt
  - (Direkte og indirekte) Objekt
  - Adverbial



## ORD – FRASER – SETNINGER

- **Setningen** - inneholder en finitt verbform og (som regel) et subjekt
  - *Spis!*
  - *Bea lukket vinduet fort.*
- **Frasen** - bygger opp setningen eller andre fraser (hierarkisk) og navngis etter hodet
  - NP (noun phrase)
  - VP (verb phrase)
  - PP (prepositional phrase)
  - etc.

- Kombinerer ord til fraser og fraser til setninger
- Beskriver forholdet mellom grupper av ord (ordklasser) og plassering i setningen
  - DT liker å komme foran et substantiv
- Syntaktiske funksjoner: subjekt, objekt
- Andre begrensninger som påvirker grammatikalitet
  - *\*Bea fant*
  - *\*Bea fant fort*
  - *\*Bea fant i bilen*
  - *Bea fant nøkkelen*

- Beskriver **hierarkisk** gruppering av ord
  - *gamle menn og kvinner kan forlate skipet*
  - [*gamle menn*] og [*kvinner*]
  - [*gamle [menn og kvinner]*]

- Strukturell flertydighet
  - *For sale: an antique desk suitable for lady with thick legs and large drawers*
  - *We will oil your sewing machine and adjust tension in your home for \$10.00*
- Flertydighet grunnet flere mulige strukturer for en setning
- Forklarer hvordan gruppering av ord relaterer til betydning

- Språkbrukere kan utføre grammatikalitetsbedømminger
- Grammatisk er ikke det samme som meningsfull
  - *Colorless green ideas sleep furiously*
  - *\*Furiously green sleep ideas colorless*
- Syntaktisk korrekt, semantisk nonsens
- Bekrefter skillet mellom syntaks og semantikk

# Syntaktisk struktur

---

- Konstituenter – grupperinger av ord i en setning, fungerer som en enhet
  - The dog ate my homework
  - The dog ate my homework
- Hvordan kan vi avgjøre konstituentstatus?
  - Lingvistiske tester

Lingvistiske tester:

- “stå alene”-testen:
  - *What did the dog eat?*
  - *My homework*
  - *\*ate my*



## Lingvistiske tester:

- “stå alene”-testen:
  - *What did the dog eat?*
  - *My homework*
  - *\*ate my*
- “erstattes med pronomener”
  - *The dog ate **it***
  - *The dog ate my homework and the cat **did** too*

## Lingvistiske tester:

- “stå alene” -testen:
  - *What did the dog eat?*
  - *My homework*
  - *\*ate my*
- “erstattes med pronomener”
  - *The dog ate **it***
  - *The dog ate my homework and the cat **did** too*
- “Flyttes som enhet”:
  - *It was **my homework** that the dog ate.*
  - ***My homework** was eaten by **the dog**.*

- Eksempel: *The dog played in the garden*

- Eksempel: *The dog played **in the garden***
  - *Where did the dog play? In the garden* (stå alene)
  - *The dog played there* (erstattes med pronomen)
  - *In the garden is where the dog played* (flytter som enhet)

**Fraser**

---

## Fraser: syntaktiske kategorier

- Et enkelt ord kan bygges ut til en gruppe ord, slik at den nye gruppen har samme funksjon i setningen (en konstituent)
- Kan substitueres for hverandre
  - *The dog ate the cake*
  - *The dog ate the birthday cake*
  - *The dog ate the delicious birthday cake*
  - *The dog ate the delicious birthday cake that was meant for Bea*

## Fraser: syntaktiske kategorier

- Et enkelt ord kan bygges ut til en gruppe ord, slik at den nye gruppen har samme funksjon i setningen (en konstituent)
- Kan substitueres for hverandre
  - *The dog ate the cake*
  - *The dog ate the birthday cake*
  - *The dog ate the delicious birthday cake*
  - *The dog ate the delicious birthday cake that was meant for Bea*
- Fraser:
  - (adledd) hode (adledd: utfylling)
  - (*the delicious birthday*) cake (*that was meant for Bea*)

# Substantivfraser (NP)

- Hodet er et substantiv
- Fungerer typisk som subjekt eller objekt i setningen
- Eksempler:
  - determinativ + substantiv: *the dog, en hund*
  - egennavn: *Barack Obama, Japan*
  - pronomen: *he, they, han, henne*
- Kongruens – feks tall, kjønn, bestemthet
- Hodet bestemmer kongruens



## Preposisjonsfraser (PP)

- Hodet er en preposisjon
- Etterfølges av en NP-utfylling (preposisjonsobjekt)
- Eksempler:
  - prep + NP *in the garden, over the rooftops*
  - foranstilt adledd (Adj) *dypest ned i skuffen*

# Adjektivfraser (AdjP)

- Hodet er et adjektiv
- Kan ha foranstilt adledd: adverb eller andre adjektiv
- Noen adjektiv tar etterstilt utfylling (som verb)
- Eksempler:
  - Adj: *crazy, red, brilliant*
  - foranstilt adledd: *almost crazy, pretty big*
  - etterstilt adledd: *crazy about dogs, lik sin far*

# Verbfraser (VP)

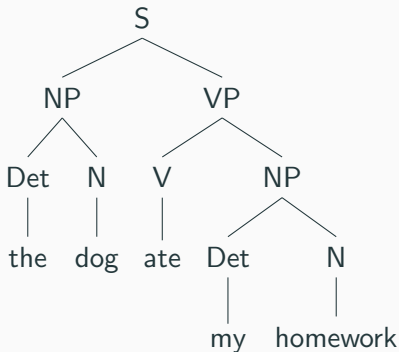
- Hodet er et verb i finitt eller infinitt form
- Eksempler:
  - verb *sover, danset*
  - verb + NP: *spiste kaken*
  - verb + NP + NP: *ga ham kaken*
  - verb + NP + PP: *la alle papirene i skuffen*
- Kan komplettere: *Jeg/Barnet/Den rare mannen . . .*

# Trær og regler

---

# Syntaktiske kategorier

- Frasale kategorier: NP, VP, AdjP, PP
- Leksikale kategorier: N, V, P, Adj, Adv
- Frasestrukturtre (Phrase Structure (PS) tree)



- *The dog ate my homework*
  - leksikale kategorier
  - finne umiddelbare konstituenter
  - finne hodet
  - ikke-leksikale kategorier (fraser)
- “top down” vs “bottom up”

1. Lineær rekkefølge i setningen
2. Syntaktiske kategorier for ord og grupper av ord
3. Hierarkisk struktur
  - **noder** som forbindes av **grener**
  - Tremetaforer: rot, løv (“leaves”), grener
  - Familiemetaforer: mor, datter, barn, søster
  - Dominans:
    - ikke-refleksiv, asymmetrisk
    - unik rot: kun en node som ikke domineres av noen annen node

- Informasjonen i et frasestrukturtre kan også representeres som **frasestrukturregler**
- Generaliserer over vår syntaktiske kunnskap
- Spesifiserer de velformede strukturene i et språk
  1.  $S \rightarrow NP VP$
  2.  $NP \rightarrow D N$
  3.  $VP \rightarrow V NP$



- Noen flere regler: intransitive verb
  1. *The cat purred*
  2. *The woman laughed*
- $VP \rightarrow V$

- Noen flere regler: PP i VP
  1. *The dog played in the garden*
  2. *The cat ate the cake on the terrace*
- VP → VP PP
- PP → P NP

- Noen flere regler: leddsetninger (innledes av subjunksjon “complementizer” (C)):
  1. *My brother said that the dog purred*
  2. *We wondered whether the cat ate the cake*
- $VP \rightarrow V CP$
- $CP \rightarrow C S$

- mengden av setninger i et naturlig språk antas å være uendelig
- språkets **kreativitet**
  - Dette er en setning
  - Dette er en setning som jeg skriver akkurat nå
  - Dette er en setning som jeg tror at jeg skriver akkurat nå
  - Dette er en setning som Fredrik mener at jeg tror at jeg skriver akkurat nå
  - osv.
  
  - Dette er en kjedelig setning
  - Dette er en kjedelig kjedelig setning
  - Dette er en kjedelig kjedelig kjedelig setning
  - osv.

- endelig og eksplisitt grammatisk beskrivelse
- trenger mekanisme som kan generere (i prinsipp) uendelige strukturer
- **rekursive** strukturer: inneholder en delstruktur av samme type som helheten
  - programmeringsspråk: feks Python
  - frasestrukturgrammatikk: en trestruktur er rekursiv dersom den inneholder en node som dominerer en annen node med samme etikett

- Rekursive regler gjør at grammatikken kan generere et uendelig antall strukturer
  - *The dog played in the garden on Monday*
  - *The dog played in the garden on Monday for an hour*
  - *The dog played in the garden on Monday for an hour with a stick*
- VP → V PP ??
- VP → VP PP

- Andre rekursive regler
  - *The dog with the collar barked*
  - *The dog with the collar around its neck barked*
  - *The dog with the collar around its neck on the sofa barked*
  - ...
- NP → NP PP

- Annet eksempel på rekursjon: adjektiver
  - The kindhearted intelligent handsome boy
  - NP → Adj NP? **NEI**
- Determinativer og adjektiver har forskjellig status innenfor substantivfrasen
  - kun én determinativ, flere adjektiv
  - adjektiver modifierer substantivet direkte, determinativet har rekkevidde over Adj+N
- Introduserer et nivå mellom NP og N: **N'**
- NP → Det N'
- N' → Adj N'
- N' → N



- Alle fraser har hode – hva med setningen?
  - The dog will eat the cake
  - The dog has eaten the cake
  - The dog is eating the cake
  - The dog might eat the cake
- VP  $\Rightarrow$  Aux VP
- Spesifiserer tidsramme for hendelsen beskrevet av verbet
- Aux er hodet: alle hendelser er lokalisert i tid

- **Seleksjon:** beskriver krav visse hoder har på sine syntaktiske omgivelser
  - verb selekterer forskjellige komplementer
    - The dog found the cake (NP) vs. \*The dog found
    - My brother said that the cat ate it (CP) vs \*My brother said
    - The cat purred ( ) vs \*The cat purred my brother

# Transformasjoner

- Vi har kunnskap om at visse setninger er relatert
- Strukturelle forskjeller er helt systematisk relatert til en forskjell i betydelse
- Deklarative setninger vs. ja/nei-spørsmål
  - The cat is purring
  - Is the cat purring?
- Kan modelleres ved **transformasjonelle regler**:
  - Flytt Aux til S
- To steg:
  1. Frasestrukturreglene genererer den grunnleggende strukturen (“d(eep)-structure”)
  2. Transformasjonelle regler kan endre (“s(urface)-structure”)

- Andre eksempler:
  - aktiv-passiv
    - The dog ate the cake
    - The cake was eaten by the dog
  - eksistensialkonstruksjoner
    - The dog was in the garden
    - There was a dog in the garden
  - topikalisering
    - Vi skal treffes i morgen
    - I morgen skal vi treffes
  - utbryting (“cleft”)
    - The dog ate the cake
    - It was the dog that ate the cake

- Generativ grammatikk: språket er lokalisert i den menneskelige hjerne
- Mål: en modell av denne indre språkmodulen (UG)
- Kan brukes til å beskrive alle naturlige språk, forutsi grammatikalitet
- Språk: enkle variasjoner over samme grunnleggende design
  - Frasestruktur
  - Fraser: hoder og komplementer
  - Setningers hode er Aux (eller T(ense)) som spesifiserer tempus, bøyningssamsvar og modalitet
- Forskjeller, feks ulik leddstilling: **parametre**