

## Syntax

Språkteknologi DH2418

Ola Knutsson

knutsson@csc.kth.se

## Innehåll

- Grundläggande begrepp
- Två perspektiv på syntax
- Frasstrukturgrammatiker
- Olika frastyper och regler för dessa
- Dependensgrammatik
- Olika sätt att representera syntaktisk information
- Jämförelser med frastrukturgrammatik och dependensgrammatik

## Varför är syntax fascinerande?

- Lingvistik som vetenskap: språket som objekt eller verktyg?
- Gärdenfors tankar om en språklig evolution - syntax växer fram för att snabba upp kommunikationen.
- Bygger på “regler” som “bestäms” på nations- eller folkgruppsnivå?

## Kan allt delas upp i små delar?

- Integration av syntax och semantik
- Sammansatta ord - syntax och semantik flyttar in i ordet: *regler för språk* vs. *språkregler*
- Flerordsuttryck - uppdelning av dessa ger inte deras syntax eller semantik: *om jag vore i dina kläder, det gäller att ha is i magen, över huvud taget*

4

## Vad är syntax?

- Syntax = studiet av hur satser byggs upp av ord och kombinationer av ord (fraser)
- Grammatik = studiet av hur satser byggs upp (av fonem, morfem, ord och ordkombinationer)
- Syntax – Semantik? (*Färglösa gröna idéer sover rasande* eller *En man har en hund*)

## Olika språksyn

a generative grammar is capable of generating an infinite number of strings from a finite set of rules (ej exakt citat från Chomsky)

“an infinite system that generates only a finite body of texts” (M.A.K. Halliday)

6

## Taggmönster som en enkel syntaktisk struktur

- Min bror väntade på flygplatsen  
PS + NN + VB + PP + NN
- Min flygplatsen väntade på bror  
PS + NN + VB + PP + NN

## Taggmönster med särdrag

PS + NN.IND + VB + PP + NN.DEF

Godkänner: *Min bror väntade på flygplatsen.*

Godkänner inte: \* *Min flygplatsen väntade på bror.*

## Övning - Swetwolmeningar

- Vår grundinställning är att alla våra svenska spelbolag som hittar på nya produkter självfallet ska få tillstånd för dem under förutsättning att de uppfyller de krav på spelansvar som vi ställer, säger vice statsminister, tillika idrottsminister Bosse Ringholm.

## Grammatikalisk och acceptabel

**+gram & +accept:** *Hon åt en grön sak.*

**+gram & -accept:** *Hon åt en grön sak.*

**-gram & -accept:** *Hon åt grön saker.*

**-gram & +accept:** *?He is a not unintelligent person (\* a not grey sky)*

*“All grammars leak” (Sapir, 1921).*

### Svensk ordföljd – rätt eller fel bland 24 teoretiskt möjliga satser (obrutna fraser)

- a. Rune köpte sin nya klocka i lördags.
- b. Rune köpte i lördags sin nya klocka
- c. I lördags köpte Rune sin nya klocka
- d. Sin nya klocka köpte Rune i lördags
- e. Sin nya klocka köpte i lördags Rune.
- f. Köpte Rune sin nya klocka i lördags?
- g. Köpte Rune i lördags sin nya klocka?
- h. Köpte i lördags Rune sin nya klocka?
- i. Rune sin nya klocka köpte i lördags.
- j. Rune i lördags köpte sin nya klocka.
- k. Rune sin nya klocka köpte i lördags.
- l. Rune i lördags sin nya klocka köpte.
- m. I lördags Rune köpte sin nya klocka.
- n. I lördags Rune sin nya klocka köpte.
- o. I lördags sin nya klocka Rune köpte.
- p. I lördags sin nya klocka köpte Rune.
- q. I lördags köpte sin nya klocka.
- r. Sin nya klocka Rune köpte i lördags.
- s. Sin nya klocka Rune i lördags köpte.
- t. Sin nya klocka i lördags Rune köpte.
- u. Sin nya klocka i lördags köpte Rune.
- v. Köpte sin nya klocka i lördags Rune?
- x. Köpte i lördags sin nya klocka Rune?
- y. Köpte sin nya klocka Rune i lördags?

## Colorless green ideas sleep furiously

- Grammaticality: någon slags grammatik bestämmer om en sats tillhör språket eller inte.
- Manning (2002): form, context and meaning.
- Acceptability är svårt bygger på individuella bedömningar
- Strangeness: språk som det används av språkakrobater
- Infelicity: fel när det gäller olika talakter

## Ordföljden i svenskan och i andra språk

S=subjekt, V=verb, O=objekt

SVO (svenska, finska, engelska)

SOV (persiska, mongoliska)

VSO (iriska, zapotek)

VOS (tagalog)

OVS (hixkarayana)

## Argument för frasstrukturgrammatik

- Språkteknologiska motiv?
- Gruppering (perceptuell grundegenskap?)
- Transformation – en viktig egenskap i språket: den lille mannen på bänken --> mannen.
- Mindre lexikon – större grammatik

## Universiell grammatik

- Det enskilda språkets variation är inte gränslös.
- En beskrivning av de gemensamma dragen i alla världens språk.
- En grammatik för enskilt språk skulle bestå av dess språkspecifika egenskaper.

## Grundläggande begrepp inom konstituent/frasstrukturgrammatik

- “grammar”
- omskrivningsregler
- konstituenten
- frasstruktur
- kontextfri grammatik

## Generativ grammatik (Chomsky)

Mål: Hitta de grammatiska regler som genererar ett språk - beskriva satsers struktur.

- Produktion och tolkning av språkliga satser involverar successiva tillämpningar av omskrivningsregler.
- En talare av ett språk måste känna språkets grammatik.
- Grammatiken definierar vilka strängar som tillhör språket eller inte.

## En grammatik definieras genom:

1. Ett lexikon
2. En uppsättning kategorier
3. En uppsättning regler

Ord - terminala symboler

Kategorier - icke-terminala symboler

Kontextfria regler: C --> s

Kontextkänsliga regler: C --> s/s1\_s2

## Kontextfria regler

Icke-terminaler

Terminaler

$S \rightarrow NP VP$

S = startsymbol (kan bestå av flera S)

## Fraser och fraser som består av fraser

Enkla NP:

$NP \rightarrow DT NN$  (en bil)

$NP \rightarrow PM$  (Pelle)

Sammansatta:

$NP \rightarrow NP KN NP$  (en bil och Pelle)

## Fler vanliga frastyper

$VP \rightarrow VB NP$  (spelar Pelle)

$PP \rightarrow Prep NP$  (i en bil)

$AP \rightarrow JJ JJ$  (stora stygga)

## En generativ grammatik:

|                  |         |                  |          |
|------------------|---------|------------------|----------|
| $S \rightarrow$  | $NP VP$ | $NP \rightarrow$ | $PM$     |
| $S \rightarrow$  | $VP$    | $NP \rightarrow$ | $NN$     |
| $VP \rightarrow$ | $VB$    | $PM \rightarrow$ | kalle    |
| $VP \rightarrow$ | $VB PP$ | $NN \rightarrow$ | regnet   |
| $PP \rightarrow$ | $P NP$  | $V \rightarrow$  | springer |
|                  |         | $P \rightarrow$  | i        |

*Kalle springer i regnet.*

*Springer i regnet.*

*?Regnet springer i Kalle*

## Särdrag i grammatiker

Man vill kunna kontrollera **kongruens**:

*det lilla huset* vs. *den lilla huset*

Verbens **subkategorisering**:

*Jag gav mannen boken. Hon sov sängen.*

*Jag tog boken mannen.*

**Ordföljd?**

## Chomsky-hierarkin

Typ 0: Obegränsade grammatikor

Typ 1: Kontextkänsliga grammatikor

Typ 2: Kontextfria grammatikor

Typ 3: Finite-state grammatikor

Grammatikens språk:

- Samma mänskligt språk – svagt ekvivalenta
- Isomorfa syntaktiska strukturer – starkt ekvivalenta.

## Dependensgrammatik

- Djupa historiska rötter, antiken och indiska grammatiker. Formell grammatik för Sanskrit (ca. 300 f.kr.)
- Föll i glömska, väcktes delvis av statistiska metoder inom språkteknologin. Men även av Constraint Grammar.

## Argument för dependensgrammatik

- Språkteknologiska motiv?
- Enklare att hantera friare ordföljd
- Mer i lexikon – mindre i grammatiken
- Mindre komplexa regler? – relationer mellan ord. Inte relationer mellan ord och fraser, fraser i fraser, satser i fraser etc.
- Tydligare koppling till semantiken (hävdar en del)

## Dependensrelationer

- En dependensrelation är en asymmetrisk relation mellan ett ord som kallas huvud (head) och ett ord som kallas modifierare (modifier).
- Ett ord kan vara ett huvud i flera relationer.
- Huvudet kan ha flera modifierare.
- Varje ord kan endast vara modifierare en gång.
- Undantag: ett speciellt ord kan inte vara modifierare överhuvudtaget, detta får namnet rot (root).
- Mängden av dependensrelationer som definieras för en mening utgör ett depedensträd.

## Syntaktiska funktioner

- Subjekt
- Objekt
- Indirekt objekt
- Prepositionskomplement
- Predikativkomplement
- Olika adverbial
- Attribut (premodifierare)
- Postmodifierare (typ PP)

## Regler inom dependensgrammatik

1. Kan  $w1$  vara vänster modifierare till huvudet  $w2$  med dependensrelationen  $r$ ?  
 $w1 \leftarrow w2$
2. Kan  $w2$  vara höger modifierare till huvudet  $w1$  med dependensrelationen  $r$ ?  
 $w1 \rightarrow w2$

## Olika sätt att representera syntaktisk struktur

- Peter gillar sin glada pojke.

## En vanlig mening från Stockholm Umeå Corpus

*Viktigaste redskapen vid ympning är  
annars papper och penna, menade han.*

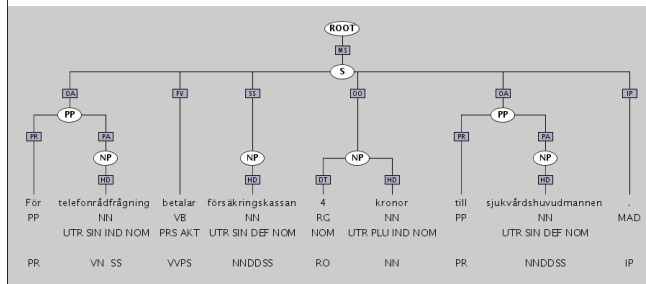
## Parseträd för frasstruktur

## Dependensträd

## Trädbanker

- en större samling syntaxträd (syntaktiskt annoterade meningar)
- träningsmaterial för maskininlärning
- utvärdering
- empiriskt sökbart material för lingvister
- man kan utvinna en grammatik för det språk som finns i korpusen
- Penn Treebank
- SynTag (100 000 ord)

### Swedish Treebank



### Swedish Treebank



Bracketing

((CL (NP (AP Viktigaste) redskapen)  
  (PP vid (NP ympning))  
  (VC är)  
  (ADVP annars)  
  (NP (NP papper) och (NP penna))))  
(CL ,  
(VC menade)  
(NP han)) . )

| # Text | Baseform          | Syntactic rel | Syntax and morphology |
|--------|-------------------|---------------|-----------------------|
| 1      | Viktigaste viktig | attr:>2       | %>N <Sup> A NOM       |
| 2      | redskapen redskap | subj:>5       | %NH N PL NOM          |
| 3      | vid vid           | mod:>2        | %N< PREP              |
| 4      | ympning ympning   | pcomp:>3      | %NH N SG NOM          |
| 5      | är vara           | obj:>11       | %MV V PRES            |
| 6      | annars annars     | advl:>5       | %AH ADV               |
| 7      | papper papper     | sc:>5         | %NH N NOM             |
| 8      | och och           | cc:>7         | %CC CC                |
| 9      | penna penna       | cc:>7         | %NH N SG NOM          |
| 10     | ,                 |               |                       |
| 11     | menade mena       | main:>0       | %MV V PAST            |
| 12     | han han           | subj:>11      | %NH PRON SG NOM       |
| 13     | .                 |               |                       |

|                                 |         |     |
|---------------------------------|---------|-----|
| Viktigaste (the most important) | APB NPB | CLB |
| redskapen (tools)               | NPI     | CLI |
| vid (in)                        | PPB     | CLI |
| ympning (grafting)              | NPB PPI | CLI |
| är (is)                         | VCB     | CLI |
| annars (normally)               | ADVPB   | CLI |
| papper (paper)                  | NPB NPB | CLI |
| och (and)                       | NPI     | CLI |
| penna (pen)                     | NPB NPI | CLI |
| ,                               | 0       | CLB |
| menade (meant)                  | VCB     | CLI |
| han (he)                        | NPB     | CLI |
| .                               | 0       | CLI |

CFG vs. DG

|                        | Kontextfri grammatik   | Dependens-grammatik |
|------------------------|------------------------|---------------------|
| Noder                  | Fras och terminaler    | Endast terminaler   |
| Syntaktiska Funktioner | Nej, endast kategorier | Ja                  |
| Ordföljd               | Ja, regler             | Nej, inga krav      |
| Lexikalisering         | Nej, regelbaserat      | Ja (valens)         |