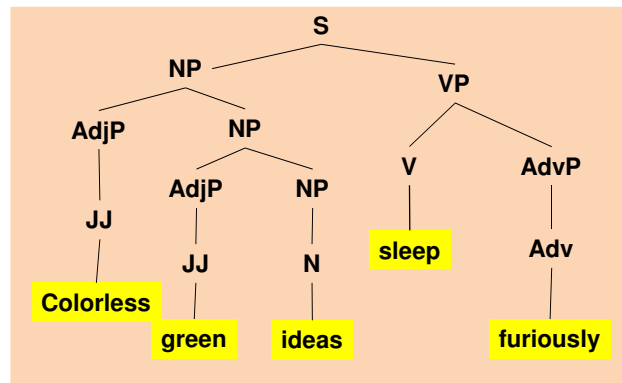


Syntax



DH2418 Språkteknologi
Johan Boye

Syntax

- Frågor vi vill besvara:
 - Vilka sekvenser av ord tillhör språket?
 - Vilka relationer finns mellan orden?
- Formella beskrivningar av ett språk:
 - Ändliga automater?
 - Grammatiker, speciellt:
 - **Kontextfri grammatik**, bygger på **konstituent**er som t.ex. nominalfras, verbfras, osv.
 - **Dependensgrammatik**, bygger på **satsdelar** (funktioner)

Tillhör språket?

- Köpte Rune sin nya klocka i lördags?
- Köpte Rune i lördags sin nya klocka?
- Köpte i lördags Rune sin nya klocka?
- Hon slog han på truten.
- Jag hörde inte låten förens igår.
- Han hade en nära döden upplevelse.
- Kalle är bättre än mig.
- Stina gick ända tills skolan.

Syntax ≠ innebörd

'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe:
All mimsy were the borogoves,
And the mome raths outgrabe.

"Beware the Jabberwock, my son!
The jaws that bite, the claws that catch!
Beware the Jubjub bird,
and shun The frumious Bandersnatch!"

He took his vorpal sword in hand:
Long time the manxome foe he sought—
So rested he by the Tumtum tree,
And stood awhile in thought.

Lewis Carroll— utdrag
ur *Jabberwocky*

Ordklasser / kategorier

- Ord delas upp i kategorier eller **ordklasser** (**POS, parts-of-speech, taggar**) efter sina morfologiska och syntaktiska egenskaper.
- T.ex. **substantiv** kan
 - förekomma efter determinerare (*ett pären*)
 - sättas i genitiv-form (*pärenets, Kalles, IBMs*)
 - sättas i plural-form (med vissa undantag, t.ex. *snö*)
- Finns ingen allmänt accepterad mängd kategorier
 - men flera standarder, t.ex. SUC (Stockholm Umeå Corpus)

Kategorier i SUC (öppna)

Kod	Kategori	Exempel
NN	Substantiv	pären
PM	Egennamn	Kalle
JJ	Adjektiv	grön
AB	Adverb	ofta
VB	Verb	springa
PC	Particip	slängd
IN	Interjektion	aj
UO	Utländskt ord	the

Kategorier i SUC (slutna)

DT	Determinerare	denna
HA	Frågande/relativt adverb	när
HD	Frågande/relativ determinerare	vilken
HP	Frågande/relativt pronomen	vem, som
HS	Frågande/relativt poss. pron.	vems, vars
IE	Infinitivmärke	att
KN	Konjunktion	och
PL	Partikel	på
PN	Pronomen	han
PP	Preposition	under
PS	Possessivt pronomen	din
RG	Grundtal	fyra
RO	Ordningstal	fjärde
SN	Subjunktion	att

Ordklasser / kategorier

- Kategorier är naturliga byggstenar i beskrivningar av språk
- Ord av samma kategori kan förekomma på samma plats i en sats (med vissa undantag):
 - "Det sitter en **fluga** på väggen"
 - "Det sitter en **rädsla** på väggen"
 - Syntaktiskt rätt även om innebörden är oklar
- Men även en **grupp av ord** kan ha samma roll:

Konstituenter

- En grupp av ord som kan sägas agera som en enhet
- T.ex. en **nominalfras**:
 - **flickan** började gråta
 - **den lilla flickan** började gråta
 - **den lilla flickan i den röda kappan** började gråta
 - **den lilla flickan i den röda kappan vars föräldrar var lärare** började gråta
- De markerade fraserna ovan kan förekomma i samma kontexter

Konstituenter

- Finns ingen allmän accepterad uppsättning konstituenter
 - precis som med kategorier
 - därför finns det många olika grammatiska teorier och alternativa formaliseringar av samma data

Kontextfria grammatiker

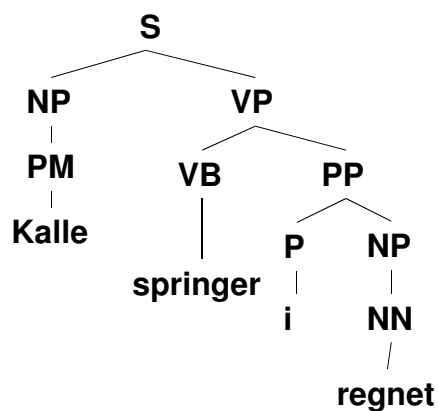
- Andra namn: **CFG**, **frasstrukturgrammatik**, **Backus-Naur Form (BNF)**
- består av:
 - terminaler (**ord**, i vårt fall)
 - icke-terminaler (**konstituenten**, i vårt fall)
 - regler på formen $A \rightarrow \alpha$, där α är en sekvens av terminaler och icke-terminaler

Kontextfria grammatiker: Exempel

$S \rightarrow NP VP$	$NP \rightarrow PM$
$S \rightarrow VP$	$NP \rightarrow NN$
$VP \rightarrow VB$	$PM \rightarrow \text{Kalle}$
$VP \rightarrow VB PP$	$NN \rightarrow \text{regnet}$
$PP \rightarrow P NP$	$V \rightarrow \text{springer}$
	$P \rightarrow i$

Kalle springer i regnet
springer i regnet
regnet springer i Kalle

Syntaxträd



[S [NP [PM Kalle]] [VP [VB springer] [PP [P i] [NP [NN regnet]]]]

Övning

- En vanlig mening från SUC-korpuset:
 - *Viktigaste redskapen vid ympning är annars papper och penna.*
 - Med kategorier: *Viktigaste (JJ) redskapen (NN) vid (PP) ympning (NN) är (VB) annars (AB) papper (NN) och (KN) penna (NN).*
- Konstruera ett syntaxträd för denna mening!
(Hitta på troliga grammatikregler)

Övning

- Utöka meningen så här:
 - *Viktigaste redskapen vid ympning är annars papper och penna, menade han.*
 - Med kategorier: *Viktigaste (JJ) redskapen (NN) vid (PP) ympning (NN) är (VB) annars (AB) papper (NN) och (KN) penna (NN) MID menade (VB) han (PN).*
- Hur skulle du bygga på syntaxträdet i föregående övning?

Parsning

- **Parsning** — processen att utifrån en mening och en grammatik konstruera ett eller flera syntaxträd.
- Varje syntaxträd svarar mot en tolkning av meningen.
- Finns många olika parsnings-algoritmer.
- Mer om detta i **föreläsning 7** (Mats W)

Kongruens

- Man vill även kunna kontrollera **kongruens**:
 - *den svarta svanen / det svart svanen*
- Inte helt okomplicerat alla gånger:
 - *How you can claim that Bolton can beat Manchester United with all their well-paided superstars surprises me.*
- För detta krävs att man håller reda på alla ords **särdrag**.

Särdrag i SUC

Särdrag	Värde	Förklaring	Kategorier
Genus	UTR	utrum	DT HD HP JJ NN PC PN PS RG RO
	NEU	neutrum	
	MAS	maskulinum	
Numerus	SIN	singular	DT HD HP JJ NN PC PN PS RG RO
	PLU	plural	
Species	IND	indefinit	DT HD HP JJ NN PC PN PS RG RO
	DEF	definit	
Kasus	NOM	nominativ	JJ NN PC PM RG RO
	GEN	genitiv	
Tempus	PRS	presens	VB
	PRT	preteritum	
	SUP	supinum	
	INF	infinitiv	

Särdrag i SUC (2)

Särdrag	Värde	Förklaring	Kategorier
Diates	AKT	aktivum	VB
	SFO	S-form	
Modus	KON	konjunktiv	VB
Particip-form	PRS	presens	PC
	PRF	perfekt	
Grad	POS	positiv	AB JJ
	KOM	komparativ	
	SUV	superlativ	
Pronomen-form	SUB	subjektsform	PN
	OBJ	objektsform	
?	SMS	sammansättn.	Alla

Exempel

Det DT NEU SIN DEF
är VB PRE AKT
ingen PN UTR SIN IND SUB/OBJ
slump NN UTR SIN IND NOM
att KN
den PN UTR SIN IND SUB/OBJ
kallas VB PRE SFO
för PP
världens NN UTR SIN DEF GEN
vackraste JJ SUV UTR/NEU SIN/PLU IND/DEF NOM
sjöresa NN UTR SIN IND NOM

Fråga:
 Vilka ord
 kongruerar
 med
 varandra?

Kongruens

- Hur kan vi koda kongruens i en kontextfri grammatik?
- $NP \rightarrow DT JJ N$ tillåter
 - *den svarta svanen*
 - **det svarta svanen*
 - **det svart svan*
 - osv

Verbens subkategorisering

- *Johan nös.* $VP \rightarrow V$
- **Johan nös en nyckel.*
- *Kalle lade en nyckel på bordet.* $VP \rightarrow V NP PP$
- **Kalle lade.*
- **Kalle lade en nyckel.*
- *Kalle gav honom ett paket.* $VP \rightarrow V NP NP$
- *Kalle gav ett paket.* $VP \rightarrow V NP$
- **Kalle gav honom.*
- *Eva påstod att han hade ljugit.* $VP \rightarrow V att S$
- **Eva påstod*
- **Eva påstod ett påstående.*

Verbens subkategorisering

- Alla verb kan inte förekomma i alla VP-regler:
 - **Johan nös en nyckel.*
 - **Kalle lade en nyckel.*
 - **Eva påstod ett påstående.*
- Vi kan subkategorisera verben efter vilka VP-regler de kan förekomma i.
 - Generalisering av transitivt/intransitivt verb
 - Moderna grammatiker kan ha tiotals eller hundratals klasser av verb.
 - Subkategorisering av verb kan hanteras på samma sätt som kongruens

Övning

- Konstruera fem VP-regler för verbet *berätta*.
- Konstruera en VP-regel som **inte** funkar för verbet *berätta*.

Alltså:

- Med kontextfria grammatiker kan vi **hantera många (alla?) syntaktiska fenomen**
- Men det finns problem:
 - t.ex. kongruens, subkategorisering
 - kan hanteras inom formalismen, men grammatikerna blir ofta stora
- Finns bättre lösningar (t.ex. unifieringsgrammatik, kap 15 i boken).
 - mer kraftfulla än kontextfria grammatiker (högre i Chomsky-hierarkin)

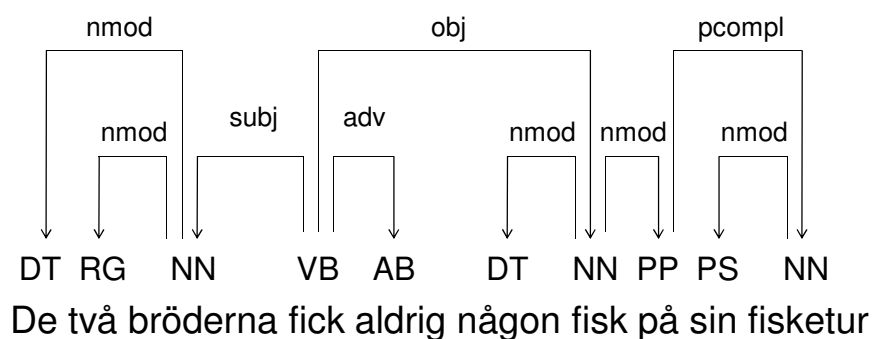
Ändliga automater?

- Räcker det inte med ändliga automater?
- Nej, pga rekursion som inte är höger-rekursion:
 - $S \rightarrow \text{if } S \text{ then } S$
 - $S \rightarrow \text{either } S \text{ or } S$
 - $S \rightarrow NP \ VP$
 - $NP \rightarrow DT \ NN$
 - $NP \rightarrow NN$
 - $VP \rightarrow VB \ NN$
 - $VP \rightarrow VB$
 - $NN \rightarrow \text{boy} \mid \text{girl} \mid \text{ice-cream} \mid \text{candy}$
 - $VB \rightarrow \text{eats}$

Dependensgrammatiker

- En **dependensgrammatik** fokuserar på ordens **funktioner** snarare än vilka konstituer de ingår i.
 - subjekt, direkt objekt, indirekt objekt, modifierare, ...
 - alltså "satsdelar"
- Utgår från (huvud-)verbet i satsen
- Binära relationer mellan ord
 - inga relationer mellan fraser och ord
- Enklare att behandla friare ordföljd

Dependensgrammatiker



Dependensgrammatiker

- En dependensrelation är en assymetrisk relation mellan ett **huvud**(-ord) och en **modifierare**.
- Ett ord kan vara huvud i flera relationer
- Ett ord kan bara vara modifierare i en relation
- Ett ord, **roten**, kan inte vara modifierare överhuvudtaget.
- Mängden av alla dependensrelationer utgör ett **dependensträd**.

Syntaktiska funktioner

- Subjekt
- Direkt objekt
- Indirekt objekt
- Prepositionskomplement
- Predikativkomplement ("predikatsfyllnad")
- Olika adverbial
- Modifierare (typ PP)

Övning

- En vanlig mening från SUC-korpuset:
 - *Viktigaste redskapen vid ympning är annars papper och penna.*
 - Med kategorier: *Viktigaste (JJ) redskapen (NN) vid (PP) ympning (NN) är (VB) annars (AB) papper (NN) och (KN) penna (NN).*
- Konstruera ett dependensträd för denna mening!

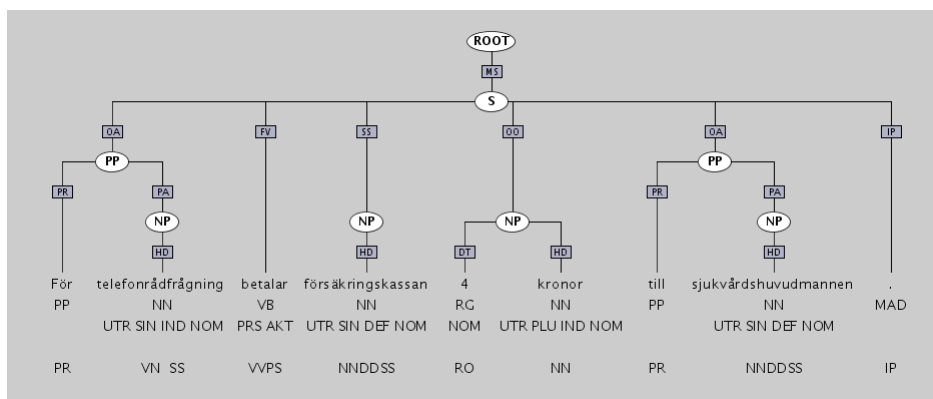
Jämförelse

	Kontextfri gram.	Dependensgram.
Noder	Konstituenten och ord	Ord
Syntaktiska funktioner	Nej	Ja
Ordföljd	Bestäms av regler	Inga krav
Lexikon	Mindre	Större
Regelmängd	Många regler	Färre regler

Trädbanker

- En **trädbank** är ett korpus där varje mening har ett associerat syntaxträd.
- Kan användas till:
 - att **inducera** fram en grammatik från exempel
 - för **utvärdering** av parsnings-algoritmer
 - som **empiriskt material** för lingvister
 - som indata för **maskininlärning**
- Penn Treebank
- Svensk Trädbank

Svensk Trädbank



Svensk Trädbank

