Björn Beskow
Torbjörn Lager
Joakim Nivre

Elementa i Generativ Grammatik

Studentlitteratur

Ev registrerade varumärken och dylikt

Innehållet i denna bok är skyddat enligt Lagen om upphovsrätt, 1960:729 och får inte reproduceras eller spridas i någon form, utan Studentlitteraturs skriftliga medgivande. Förbudet gäller hela verket såväl som delar av verket och inkluderar lagring i elektroniska media, visning på bildskärm samt bandupptagning.

Inlagan är tryckt på MultiFine Offset, licens nr 304 054

Art. nr šššššš
ISBN 91-44-šš
© šš och Studentlitteratur 19šš, 19šš
Fackgranskning: šš šš
Tecknare: šš šš
Omslagsbild: šš šš
Omslagslayout: šš šš

Printed in Sweden
Studentlitteratur, Lund

Tryckning:	1	2	3	4	5	6	7	8	9	10	År	1999	98	97	96	95

Förord

Den generativa grammatiken har sedan den introducerades för ett fyrtiotal år sedan genomgått en omfattande utveckling. Nya teorier har sett dagens ljus och gamla teorier har dött eller förvandlats till oigenkännlighet. Samtidigt har man emellertid byggt upp ett stort antal begrepp och metoder som blivit allmängods inom det generativa paradigmet och som i dag tillhör alla teorier lika väl som ingen.

En del av detta allmängods utgörs av grundläggande teoretiska antaganden (t. ex. distinktionen mellan kompetens och performans), vilka brukar få ett ganska väl tilltaget utrymme i introduktionsböcker. En minst lika stor del utgörs emellertid av mera praktiska verktyg (t. ex. frasstrukturregler och analysträd), vilka för den oerfarne studenten är minst lika oumbärliga men som tyvärr ofta behandlas mera styvmoderligt i läroböckerna.

Denna bok har tillkommit som ett försök att avhjälpa denna brist och syftet är i första hand att åstadkomma en praktisk, hantverksmässig orientering om den generativa grammatikens grundbegrepp utan alltför djupgående resonemang om de teoretiska motiveringarna bakom begreppen. Detta innebär att boken kan ses som ett komplement till existerande introduktionsböcker i generativ grammatik, där de teoretiska resonemangen ges större utrymme.

Vi har valt att centrera framställningen kring den klassiska *frasstrukturgrammatiken*, en grammatikmodell som i sin rena form inte spelar någon framträdande roll i de moderna grammatiska teorierna, men som ligger till grund för de senaste fyrtio årens utveckling, och som utan tvekan tillhör den bakgrundskunskap man måste ha för att fullt ut kunna tillgodogöra sig den nyare litteraturen.

Boken bygger på ett kompendium, som ursprungligen skrevs av Joakim Nivre, och som under ett flertal år använts i grundutbildningen i allmän språkvetenskap på Göteborgs universitet. Vår erfarenhet är att kompendiet

© Studentlitteratur

fungerat väl i undervisningen och haft en lagom omfattning för det första introducerande momentet i generativ grammatik.

På senare år har kompendiet också använts tillsammans med ett datorprogram, *Phrase Structure Grammar Laboratory (PSG Lab)*, utvecklat av författarna för laborationer och övningar i frasstrukturgrammatik och särskilt anpassat till kompendiets framställning. Vår tanke är att boken och programmet skall användas tillsammans som ett paket, och för att underlätta detta har vi som appendix inkluderat dels en handledning till programmet, dels en laboration som illustrerar hur man kan använda programmet i praktiska övningar. (Tillsammans med handledningen finns också information om hur man skaffar programmet och vad som krävs för att kunna köra det.)

I ”undervisningspaketet” ingår också de övningar som återfinns i slutet av de flesta kapitel i boken. Till dessa hör ett facit som inte bara ger lösningsförslag utan även förklarar en del nya begrepp som introduceras i övningarna. Det kan i detta sammanhang vara värt att påpeka att en del av de grammatiska analyser som presenteras, både i facit och i boken i övrigt, är medvetet förenklade för att undvika alltför långtgående teoretiska resonemang och hålla fokus på de elementära grundbegreppen. Vårt syfte har i första hand varit att skriva en pedagogisk introduktion till generativ grammatik – inte att åstadkomma en fullödig beskrivning av svensk grammatik.

Till sist vill vi rikta ett tack till alla lärare och studenter som använt det tidigare kompendiet och PSG Lab på olika kurser och vars värdefulla synpunkter bidragit till den slutliga utformningen av såväl programmet som denna bok. De brister som ändå finns kvar står författarna själva för.

Göteborg, januari 1996

Författarna

© Studentlitteratur

Förord 29 January 1996 21.23:04 sida iv av iv

Innehåll

Generativ grammatik 7
Det generativa paradigmet 7
Språk och grammatik 8
Satser 8
Grammatiska beskrivningar 10
Universell grammatik 13
Teori och data 15
Grammatik och syntax 16
Satsers struktur 19
Linjär struktur 19
Konstituentstruktur 21
Övningar 25
Syntaktiska kategorier 27
Intrinsikala och relationella kategorier 28
Lexikala och icke-lexikala kategorier 29
Övningar 32
Analysträd 33
Noder och grenar 33
Etiketterade noder 35
Precedensrelationer i analysträd 36
Övningar 38
Frasstrukturgrammatik 39
Frasstrukturregler 40
Strängderivation 40
Trädkonstruktion 42
Lexikon 43
En enkel frasstrukturgrammatisk modell 46
Ett fragment av svenska 49

© Studentlitteratur

Syntaktiska kategorisymboler 49
Lexikon 50
Frasstrukturregler 51
Exempel 54
Övningar 55
Mer om frasstrukturgrammatik 57
Förkortningskonventioner 57
Rekursiva strukturer 61
Övningar 65
Bortom frasstrukturgrammatik 67
Problem med frasstrukturgrammatiska beskrivningar 67
Frasstrukturgrammatikens roll i modern grammatisk teori 70

© Studentlitteratur

elementaIH 29 januari 1996 20.56:42 sida 6 av 6

1 Generativ grammatik

1.1 Det generativa paradigmet

Termen *generativ grammatik* används här som en samlande benämning på den typ av grammatiska teorier som varit dominerande under de senaste trettio åren av lingvistikens historia. Grunden till den generativa grammatiken lades på 1950-talet av den amerikanske lingvisten Noam Chomsky, och de första decennierna av dess historia dominerades helt av den *transformationella* grammatikteori som Chomsky utvecklade bl. a. i de klassiska verken *Syntactic Structures* (1957) och *Aspects of the Theory of Syntax* (1965).

De senaste tjugo åren har emellertid sett ett flertal alternativa teorier växa fram vid sidan av den utveckling som Chomskys teorier genomgått. Dessa teorier skiljer sig från varandra på många sätt men delar samtidigt ett antal grundläggande antaganden av teoretisk och metodologisk natur, tillräckligt många för att det enligt vår mening skall vara befogat att tala om ett generativt paradigm inom lingvistiken.

Bland de teorier som varit mest tongivande under det senaste decenniet kan nämnas *Government and Binding Theory* (GB) och dess efterföljare *Minimalistprogrammet*, vars främsta företrädare är Noam Chomsky; *Generalized Phrase Structure Grammar* (GPSG), som utvecklats framför allt av Gerald Gazdar; *Lexical-Functional Grammar* (LFG), som i första hand förknippas med Joan Bresnan; samt *Head-Driven Phrase Structure Grammar*, som lanserats av Carl Pollard och Ivan Sag.

Dessa teorier delar som sagt ett flertal grundläggande teoretiska antaganden om språk och grammatik, men de delar också en grundläggande begreppsapparat av mera praktisk och hantverksmässig natur, en metodologisk och terminologisk tradition som man måste vara insatt i för att kunna tillgodogöra sig teorierna.

Det främsta syftet med denna bok är att ge en introduktion till denna tradition och att förmedla den generativa grammatikens praktiska grundbegrepp.

© Studentlitteratur

Tonvikten ligger alltså på den praktiska sidan, men vi vill ändå börja med att kortfattat ta upp några av de mest centrala teoretiska antaganden som kännetecknar det generativa paradigmet.

1.2 Språk och grammatik

Vad är ett språk? Frågan kan besvaras på många olika sätt beroende på vilket perspektiv vi utgår ifrån. Vi kan t. ex. beskriva ett språk som ett teckensystem, som ett kommunikationsmedel, eller som ett socialt normsystem. Inom den generativa grammatiken utgår man vanligen från antagandet att ett *språk* kan betraktas som en *mängd*, närmare bestämt som en mängd av *satser*. Svenska är enligt detta synsätt en mängd som bl. a. innehåller följande satser:

(1) *min bror köpte cykeln av en gammal vän*
den gamla damen träffade killen med handväskan

En *grammatisk beskrivning* av ett språk (eller kort och gott en *grammatik*) är enligt det generativa synsättet en explicit redogörelse för den mängd av satser som utgör språket i fråga.

Dessa preliminära definitioner av begreppet *språk* som en mängd av satser och begreppet *grammatik* som en explicit redogörelse för denna mängd av satser ger dock omedelbart upphov till två nya frågor: Vad är en sats? Vad är en explicit redogörelse för en mängd av satser?

1.3 Satser

Låt oss börja med att försöka kasta lite ljus över begreppet *sats*, så som det används inom den generativa grammatiken. Som en första provisorisk karakteristik kan vi säga att en sats består av en *sekvens av ord*. Till exempel är (2) en sats som innehåller de åtta orden *min*, *bror*, *köpte*, *cykeln*, *av*, *en*, *gammal* och *vän* i en bestämd ordningsföljd.

(2) *min bror köpte cykeln av en gammal vän*

© Studentlitteratur

Kapitel 1 29 januari 1996 20.59:43 sida 8 av 18

Dessa åtta ord kan teoretiskt kombineras på 40 320 olika sätt. Av alla dessa kombinationer är det emellertid bara ett litet fåtal (kanske tio eller tolv) som är gångbara som svenska satser. Jämför t. ex. (3) och (4).

(3) *cykeln köpte min bror av en gammal vän*
köpte min bror cykeln av en gammal vän

(4) *vän gammal en av cykeln köpte bror min*
min gammal bror av en cykeln vän köpte

Alla som talar svenska kan utan problem avgöra att sekvenserna i (3) i någon mening utgör ”riktiga” satser, medan sekvenserna i (4) inte gör det (de är helt enkelt bara sekvenser av ord). I den gängse terminologin säger man att sekvenserna i (3) är *grammatikaliska* (eller *välformade*) *satser* i svenska, medan sekvenserna i (4) är *ogrammatikaliska* (*ej välformade*).[1] Om man vill markera att en viss ordsekvens är ogrammatikalisk låter man den vanligen föregås av en asterisk (*). Om vi t. ex. vill ange att ordsekvenserna i (4) inte är välformade svenska satser, skriver vi alltså på följande sätt:

(5) **vän gammal en av cykeln köpte bror min*
**min gammal bror av en cykeln vän köpte*

Det är viktigt att lägga märke till att begreppet *sats* (enligt den gängse uppfattningen bland generativa grammatiker) inte är detsamma som *acceptabelt språkligt yttrande* eller *faktiskt förekommande yttrande* eller något liknande. När vi använder språket för att kommunicera yttrar vi ofta ordsekvenser som ur den generativa grammatikens perspektiv inte motsvarar grammatikaliska satser. Således kan t. ex. (6) fungera som ett fullt begripligt yttrande om vem som har uppfört en viss byggnad, fastän de flesta torde vara överens om att det inte utgör en grammatikalisk sats.

(6) *den stora hus byggt av Kalle*

Omvänt är många grammatikaliska satser svårbegripliga och olämpliga i vardagliga kommunikationssammanhang. Betrakta t. ex. följande satser:

(7) *osten möglade*

[1] Uttrycket *grammatikalisk sats* är närmast en tautologi, eftersom endast grammatikaliska (välformade) ordsekvenser är satser över huvud taget. På motsvarande sätt är uttrycket *ogrammatikalisk sats* egentligen en motsägelse. Dessa båda uttryck används emellertid ofta i betydelsen *grammatikalisk ordsekvens* (dvs. *sats*) respektive *ogrammatikalisk ordsekvens* (dvs. *icke-sats*). Så kommer även vi att använda dem i den följande framställningen.

© Studentlitteratur

Kapitel 1 29 januari 1996 20.59:43 sida 9 av 18

osten råttan tog möglade
osten råttan katten jagade tog möglade
osten råttan katten hunden bet jagade tog möglade

Satser bildade på detta sätt, genom upprepad inbäddning av relativsatser, blir snabbt omöjliga att förstå utan papper och penna, trots att åtminstone de flesta grammatiker anser att de är fullt grammatikaliska.

Satser är således tämligen abstrakta ting, som i varje fall inte kan likställas med observerade språkliga yttranden. Uppfattningarna går isär om vad satser egentligen är för slags ting, men detta är en fråga som vi lyckligtvis inte behöver besvara här. Två saker måste vi dock notera i sammanhanget. För det första: begreppet *sats* är ett abstrakt teoretiskt begrepp som den generativa grammatiken antar i sin språkbeskrivning, inte något som man direkt kan observera i samtal. För det andra: den generativa grammatiken utgår från antagandet att den som kan ett språk har förmågan att (i de flesta fall) avgöra om en given sekvens av ord är en (grammatikalisk) sats i språket eller inte.

1.4 Grammatiska beskrivningar

I avdelning 1.2 hävdade vi att en *grammatisk beskrivning* av ett språk (eller kort och gott en *grammatik*) enligt det generativa synsättet är en explicit redogörelse för den mängd av satser som utgör språket i fråga. Låt oss nu ta upp frågan om vad som menas med en *explicit redogörelse för en mängd av satser*.

Det första kravet en sådan redogörelse måste uppfylla är att exakt definiera den aktuella mängden av satser, dvs. att tala om precis vilka sekvenser av ord som ingår i mängden och vilka som inte gör det. En korrekt grammatisk beskrivning av svenska måste alltså på något sätt tala om att sekvenserna i (8) ingår i mängden av svenska satser medan sekvenserna i (9) inte gör det.

(8) *min bror köpte cykeln av en gammal vän*
den gamla damen träffade killen med handväskan

(9) **vän gammal en av cykeln köpte bror min*
**handväskan med killen träffade damen gamla den*

© Studentlitteratur

Om en grammatisk beskrivning uppfyller detta krav, dvs. om den skiljer alla satser från icke-satser i ett visst språk, säger man att grammatiken *genererar alla och endast språkets grammatikaliska satser* (dvs. den genererar alla grammatikaliska satser i språket och den genererar endast grammatikaliska satser). Härav termen *generativ* grammatik som i detta sammanhang alltså inte betyder mer än ”exakt definierande”.[2]

Hur skall man då utforma en generativ beskrivning av mängden satser i ett språk? Allmänt sett finns det två olika sätt att ange en mängd (vare sig det är en mängd av satser eller en mängd av andra ting). Det första är att helt enkelt räkna upp mängdens medlemmar. Det andra är att ge en beskrivning som passar in på exakt de element som ingår i mängden och inga andra. I (10–11) exemplifieras hur en och samma mängd kan anges på två olika sätt:[3]

(10) { Helsingfors, Köpenhamn, Oslo, Reykjavik, Stockholm }

(11) { x | x är huvudstad i ett nordiskt land }

I (10) anger vi mängden av nordiska huvudstäder *extensionellt* (genom *uppräkning*); i (11) anger vi den *intensionellt* (genom *beskrivning*).

Att ge en *extensionell* redogörelse för mängden av satser i ett språk skulle alltså innebära att man räknade upp samtliga satser i en lång lista. De flesta är emellertid överens om att en sådan redogörelse vore olämplig som grammatisk beskrivning av flera skäl.

För det första brukar man anta att varje naturligt språk består av ett oändligt antal satser, vilket alltså skulle medföra att en eventuell lista aldrig skulle bli fullständig hur länge vi än höll på.[4]

Ett annat skäl har att göra med ett ytterligare krav som man inom det generativa paradigmet ställer på en grammatisk beskrivning förutom att den skall generera alla och endast de grammatikaliska satserna i ett språk, nämligen att den skall redogöra för satsernas *struktur*. En korrekt grammatisk beskrivning av svenska måste enligt detta synsätt tala om inte bara att (12) är en välformad (grammatikalisk) sats på svenska, utan också att den primärt består av delarna *min bror* och *köpte cykeln av en gammal vän*, där

[2] En grammatik som genererar alla och endast de grammatikaliska satserna i ett språk sägs ibland vara *observationsmässigt adekvat*. En varning är dock på sin plats eftersom denna term också används på andra sätt.

[3] För en introduktion till mängdlära och mängdteoretisk notation, se t. ex. Allwood, Andersson & Dahl (1977) *Logic in Linguistics*, Cambridge University Press.

© Studentlitteratur

den senare delen i sin tur är uppbyggd av delarna *köpte*, *cykeln* och *av en gammal vän*, osv. ända ned till de enskilda orden som ingår i satsen.

(12) *min bror köpte cykeln av en gammal vän*

Detta innebär att en extensionell redogörelse för mängden av satser i form av en lista på alla grammatikaliska ordsekvenser skulle vara inadekvat som grammatisk beskrivning, även om den kunde göras fullständig.

Vi skall ägna oss mer åt satsers struktur i kapitel 2, men tills vidare kan vi nöja oss med att konstatera att vi nu vaskat fram två krav som man brukar ställa på en grammatisk beskrivning inom det generativa paradigmet. Det första kravet är att beskrivningen skall *generera alla och endast de grammatikaliska satserna i det aktuella språket*, dvs. den skall exakt definiera mängden av grammatikaliska satser. Det andra kravet är att beskrivningen skall *tilldela varje sats en strukturell beskrivning*, dvs. den skall redogöra för strukturen hos varje grammatikalisk sats.

En grammatik som uppfyller det första kravet sägs ibland vara *observationsmässigt adekvat*, medan en grammatik som (också) uppfyller det andra kravet sägs vara *deskriptivt adekvat*.[5] Det är värt att notera att ingen ännu lyckats formulera en fullständig, observationsmässigt och deskriptivt adekvat generativ grammatik för något naturligt språk.

Vi har också konstaterat att en (generativ) grammatisk redogörelse för ett språk, dvs. för en mängd av satser, av olika skäl måste vara en *intensionell* redogörelse, vilket alltså innebär att vi redogör för mängden av satser genom att beskriva de utmärkande egenskaperna hos en grammatikalisk sats, snarare än genom att räkna upp alla satserna i en lista (jfr 10–11 ovan).

---

[4] Standardargumentet för att antalet satser i ett naturligt språk är oändligt återfinns i de flesta introduktionsböcker till generativ grammatik och går ut på att det inte finns någon principiell gräns för hur lång en sats kan vara, dvs. oavsett hur lång en sats är, kan vi alltid bilda en längre sats t. ex. genom samordning. Givet *björnen sover* kan vi bilda *björnen sover och vargen ylar*; givet den senare satsen kan vi bilda *björnen sover och vargen ylar och månen skiner*; osv. Ett annat sätt att ”förlänga” satser i all oändlighet exemplifieras i (7) ovan. När man talar om att satser kan vara hur långa som helst är det återigen viktigt att hålla i minnet att *sats* är ett abstrakt teoretiskt begrepp som inte kan likställas med *möjligt yttrande*. Språkliga yttranden kan knappast vara hur långa som helst, eftersom talande människor måste äta och sova ibland och dessutom har en begränsad livslängd. Denna typ av begränsningar är dock (enligt standardargumentet) inte språkliga, eftersom de begränsar alla former av mänskliga aktiviteter, och har därför ingen relevans för det teoretiska begreppet *sats*.

[5] En varning är återigen på sin plats, eftersom dessa båda termer också används på andra sätt i litteraturen.

© Studentlitteratur

Kapitel 1 29 januari 1996 20.59:43 sida 12 av 18

Hur ser då ”en observationsmässigt och deskriptivt adekvat intensionell redogörelse för en mängd av satser” ut? Detta är en av huvudfrågorna inom den generativa grammatiken, och de olika konkurrerande teorier som existerar på området kan ses som olika försök att besvara denna fråga. Utan att gå in på några detaljer kan vi dock säga att den traditionella utgångspunkten är att en grammatisk beskrivning bör bestå av två huvudkomponenter: ett *lexikon* (dvs. en lista över de minsta grammatiska byggstenarna i språket, t. ex. ord eller morfem) och en uppsättning *grammatiska regler* (dvs. regler som talar om hur de minsta byggstenarna kan kombineras till större enheter och slutligen till satser). Utifrån denna utgångspunkt har sedan olika personer formulerat olika teorier om vad som bör ingå i lexikonet och om hur grammatiska regler bör se ut (mer om detta i kapitel 5–8).

1.5 Universell grammatik

Kraven på observationsmässig och deskriptiv adekvans är krav på grammatiska beskrivningar av enskilda språk, såsom svenska, ryska och japanska. Inom den generativa grammatiken har man emellertid också som ett viktigt mål att försöka åstadkomma en *universell grammatik*, dvs. att beskriva vad som är gemensamt och utmärkande för alla naturliga språk och att formulera en teori om den grammatiska strukturen hos naturliga språk i allmänhet, varav den grammatiska strukturen i specifika språk som svenska och japanska kan ses som specialfall.

Grammatisk struktur varierar mycket från språk till språk, men de flesta som arbetar inom det generativa paradigmet antar ändå att denna variation inte är gränslös, utan att det finns egenskaper som är gemensamma för alla mänskliga språk och som begränsar variationsrymden för hur en grammatik för ett mänskligt språk kan se ut.

Om vi hade en teori som beskrev de drag som är gemensamma för alla naturliga språk, skulle den grammatiska beskrivningen av enskilda språk kunna begränsas till att redogöra för de egenskaper som är språkspecifika och som varierar från språk till språk. Om vi hade en sådan teori skulle vi förmodligen också ha en metod för att avgöra vilken av två föreslagna grammatiska beskrivningar av t. ex. ryska som är den mest korrekta (därför att den stämmer bäst överens med den universella grammatiken), även om

© Studentlitteratur

Kapitel 1 29 januari 1996 20.59:43 sida 13 av 18

båda beskrivningarna i övrigt är såväl observationellt som deskriptivt adekvata beskrivningar av ryska.[6]

Om vi hade en sådan universell grammatik, skulle vi enligt många ha en *explanatoriskt adekvat* grammatisk teori.[7] En explanatoriskt adekvat teori är en teori som definierar begreppet ”möjlig grammatik för ett naturligt språk” och inom vars ram vi kan formulera grammatiska beskrivningar för enskilda naturliga språk utan att för varje språk behöva upprepa de delar av grammatiken som är universella. En sådan teori sägs vara explanatorisk, därför att den i någon mening förklarar varför enskilda språks grammatiker ser ut som de gör.

Därmed kan vi preliminärt sammanfatta några av den generativa grammatikens mål och antaganden.

Antaganden:

1. Ett *språk* är en mängd av satser.
2. En *sats* är en sekvens av ord med en bestämd struktur.
3. En (generativ) *grammatisk beskrivning* av ett språk (en *grammatik*) är en explicit redogörelse för mängden av satser i språket.

Mål:

1. Att åstadkomma korrekta grammatiska beskrivningar av enskilda naturliga språk.
2. Att åstadkomma en universell grammatisk teori, som definierar begreppet ”möjlig grammatik för ett naturligt språk”.

---

[6] Med detta menas *inte* att den universella teorin ger oss en automatisk metod för att *upptäcka* deskriptivt adekvata grammatiska beskrivningar för naturliga språk, utan bara att den hjälper oss att välja mellan redan existerande grammatiska beskrivningar.

[7] Observera att kravet på *explanatorisk adekvans* i första hand är ett krav på den universella grammatiska teorin, medan kraven på observationsmässig och deskriptiv adekvans primärt gäller grammatiska beskrivningar av enskilda språk (som enligt det generativa paradigmet helst bör vara formulerade inom ramen för en universell grammatisk teori).

© Studentlitteratur

Kapitel 1 29 januari 1996 20.59:43 sida 14 av 18

1.6 Teori och data

Vi har hittills i detta kapitel behandlat några av den generativa grammatikens teoretiska antaganden om språk och grammatik. Till de grundantaganden som kännetecknar det generativa paradigmet hör emellertid också antaganden av mera allmän metodologisk och vetenskapsteoretisk natur, och vi vill därför gärna säga några ord om den generativa grammatikens metodologiska grundsyn.

Som tidigare nämnts (avdelning 1.3) utgår man inom det generativa paradigmet från antagandet att den som kan ett språk har förmågan att (i de flesta fall) avgöra om en given sekvens av ord är en sats i språket eller ej. Man antar vidare att dessa intuitiva omdömen[8] utgör den bästa källan till kunskap om språket, bättre än t. ex. empiriska undersökningar av samtal och andra typer av konkret språkanvändning.

Intuitiva omdömen av detta slag sägs därför ofta utgöra den generativa grammatikens *data*, dvs. de utgör grunden för de teorier man ställer upp om ett språks grammatik. Utifrån dessa data formuleras *hypoteser*, vars förutsägelser kan testas mot nya data (dvs. nya intuitiva omdömen). Denna metod, som ibland kallas *hypotetisk-deduktiv*,[9] kan åskådliggöras med ett enkelt exempel.

Antag att vi som talare av svenska konstaterat – genom intuitiva grammatikalitetsomdömen – att ordsekvenserna i (13) utgör grammatikaliska satser i svenska och att ordsekvenserna i (14) inte gör det.

(13) *han köpte inte cykeln*
köpte han inte cykeln
cykeln köpte han inte

(14) **inte cykeln han köpte*
**inte han cykeln köpte*
**inte köpte cykeln han*

[8] Man säger att sådana omdömen är *intuitiva* därför att de antas vara baserade på språkliga *intuitioner*, som den normale språkbrukaren besitter i kraft av sin implicita kunskap om språkets grammatik. Termen *intuitiv* indikerar således i första hand att omdömena bygger på outtalad kunskap, snarare än på t. ex. objektiva mätinstrument eller statistiska undersökningar.
[9] *Hypotetisk* därför att man formulerar hypoteser, *deduktiv* därför att man sedan deducerar (härleder) testbara konsekvenser från hypoteserna.

© Studentlitteratur

De data vi har att utgå från är då att ordsekvenserna i (13) är satser, medan ordsekvenserna i (14) inte är det. Utifrån dessa data skulle vi t. ex. kunna formulera den (föga lyckade) hypotesen att alla ordsekvenser vars första element är *inte* är ogrammatikaliska på svenska.

Nästa steg blir att härleda en testbar konsekvens av denna hypotes, dvs. något som borde gälla om hypotesen var korrekt och som vi kan testa mot nya data. En sådan konsekvens är t. ex. att sekvensen *inte köpte han cykeln* borde vara ogrammatikalisk eftersom dess första element är *inte*.

Nästa steg igen blir att testa denna konsekvens mot data, dvs. mot våra intuitiva omdömen om grammatikalitet. Resultatet av detta test blir i vårt fall att konsekvensen inte gäller, eftersom våra intuitioner (eller i varje fall författarnas) säger oss att ordsekvensen *inte köpte han cykeln* utgör en alldeles utmärkt grammatikalisk sats på svenska. Detta leder i sin tur till att vi får förkasta hypotesen, eftersom den inte höll måttet (dvs. hade falska konsekvenser) i ljuset av nya data.

Efter att ha förkastat hypotesen är vi på sätt och vis tillbaka där vi startade. Men det finns en viktig skillnad. För det första har vi eliminerat en möjlig hypotes, som vi därför inte behöver testa igen. För det andra har vi utökat vår datamängd, som nu också innehåller det faktum att *inte köpte han cykeln* är en grammatikalisk sats på svenska. Med utgångspunkt i denna utvidgade datamängd kan vi så formulera nya hypoteser, vars konsekvenser kan testas mot nya data, osv.

1.7 Grammatik och syntax

Låt oss avsluta kapitlet med en terminologisk anmärkning. Vi har hittills genomgående använt termen *grammatik* när vi har talat om generativa beskrivningar av språk, betraktade som mängder av satser. En annan term som ofta förekommer i detta sammanhang är *syntax*, och det är därför motiverat att fråga hur dessa båda termer är relaterade. Problemet är emellertid att termerna *syntax* och *grammatik* används på olika sätt av olika människor. Ofta har *syntax* en snävare betydelse och betecknar då en del av grammatiken, som dessutom antas omfatta *morfologi* samt eventuellt *fonologi* (detta torde vara den mest traditionella användningen av termerna). Men ibland används termerna mer eller mindre som synonymer i den vidare bemärkelsen. Detta är speciellt vanligt inom det generativa paradigmet, där

© Studentlitteratur

Kapitel 1 29 januari 1996 20.59:43 sida 16 av 18

man ofta finner termen *generativ syntax* använd som synonym till *generativ grammatik*.

I denna bok skall vi försöka hålla oss till vad vi tror är den vanligaste och mest funktionella användningen av termerna, nämligen enligt följande:

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)

På detta sätt blir syntaxen en del av grammatiken, nämligen den del som handlar om hur ord (och större enheter) bygger upp satser, medan grammatiken som helhet handlar om hur satser byggs upp mera allmänt och därför också innefattar studiet av satsers morfologiska och fonologiska struktur.

Den generativa grammatiken omfattar därför i princip såväl fonologi och morfologi som syntax. De flesta generativa grammatiker har emellertid intresserat sig främst för syntaxen (dvs. de har antagit *ordet* som den minsta byggstenen i sina grammatiska analyser), och det torde vara på syntaxens område som den generativa grammatiken nått sina största framgångar. Som en följd av detta kommer också denna bok främst att handla om generativ syntax.

© Studentlitteratur

© Studentlitteratur

Kapitel 1 29 januari 1996 20.59:43 sida 18 av 18

2 Satsers struktur

2.1 Linjär struktur

Grammatisk teori handlar om satsers struktur. En sats kan sägas ha struktur på flera olika nivåer. Satsen *min bror köpte cykeln* kan t. ex. för det första betraktas som en sekvens av *fonem* (eller *grafem*),[1]

(15) */ minbru:rçøptesykeln /*

för det andra som en sekvens av *morfem*,

(16) {*min*} {*bror*} {*köp*} {*-te*} {*cykel*} {*-n*}

och för det tredje som en sekvens av *ord*.

(17) *min bror köpte cykeln*

I alla tre fallen analyseras satsen som en sekvens av språkliga uttryckselement (fonem, morfem respektive ord). Sådana sekvenser kallas ofta för *strängar* (fonemsträngar, morfemsträngar respektive ordsträngar). Strukturen hos en sträng sägs vara strikt *linjär*, eftersom de element som ingår i strängen kommer i en bestämd ordningsföljd. Närmare bestämt kan vi säga att elementen i en sträng är ordnade av relationen ”kommer före”, vilket innebär att om vi betraktar två olika element *x* och *y* (vilka som helst) i strängen, så gäller antingen att *x kommer före y* eller att *y kommer före x* (ingen annan möjlighet gives).[2]

---

[1] Eftersom de flesta grammatiker betraktar det talade språket som primärt i förhållande till det skrivna språket talar man hellre om *fonem* (språkljudstyper) än om *grafem* (bokstavstyper) när man diskuterar språkets minsta betydelseskiljande element; detta trots att man i praktiken mest sysslar med skrivna representationer av språk. Vi kommer här att okritiskt ansluta oss till denna tradition.

[2] Följande två fall är alltså uteslutna: (i) *x* kommer före *y och y* kommer före *x* (dvs. både-och); (ii) det är *inte* så att *x* kommer före *y och* det är *inte* så att *y* kommer före *x* (dvs. varken-eller).

© Studentlitteratur

I (17) ovan är t. ex. elementen *köpte*, *min*, *cykeln* och *bror* ordnade på följande sätt: *min* kommer före *bror*, *köpte* och *cykeln*; *bror* kommer före *köpte* och *cykeln*; och *köpte* kommer före *cykeln*. Relationen ”kommer före” kan i samband med språkliga uttryck tänkas motsvara antingen den temporala ordningsföljd som råder mellan element i talet, eller den vänster-höger-ordning som (i vårt västerländska skriftsystem) råder mellan element i skriften.

En annan relation som råder mellan element i en sträng är relationen ”kommer omedelbart före”. Denna relation råder emellertid inte mellan alla par av element, utan bara mellan de element som är närmaste grannar i strängen. I (16) ovan gäller t. ex. följande: {*min*} kommer omedelbart före {*bror*}; {*bror*} kommer omedelbart före {*köp*}; {*köp*} kommer omedelbart före {*-te*}; {*-te*} kommer omedelbart före {*cykel*}; och {*cykel*} kommer omedelbart före {*-n*}. Däremot råder inte relationen alls mellan t. ex. {*min*} och {*cykel*} (dvs. det är inte så att {*min*} kommer omedelbart före {*cykel*}, men det är inte heller så att {*cykel*} kommer omedelbart före {*min*}).[3]

Satser kan alltså betraktas som fonemsträngar, morfemsträngar eller ordsträngar. Inom den generativa grammatiken har man emellertid ofta utgått från ordet som den minsta grammatiska byggstenen, dvs. man har i första hand intresserat sig för satsers syntaktiska struktur (och i stor utsträckning lämnat den morfologiska och fonologiska strukturen därhän). Detta innebär att den enda typ av strängar man i praktiken intresserat sig för är ordsträngar. I den följande framställningen kommer vi i stort att ansluta oss till denna praxis, men det kan vara värt att hålla i minnet att satser också kan ses som strängar av fonem eller morfem.

Innan vi går vidare skall vi uppmärksamma ytterligare en egenskap hos strängar. Om man sätter en sträng omedelbart före en annan sträng, så blir resultatet en ny sträng. Operationen kallas *konkatenering* (lat. ”sammanlänkning”) och illustreras i (18–20), där (20) kan ses som resultatet av att konkatenera (18) med (19).

---

[3] För relationen ”kommer före” gäller följande: (i) inget element kommer före sig själv i en sträng; (ii) om *x* kommer före *y* i en sträng, så kommer inte *y* före *x*; (iii) om *x* kommer före *y* och *y* kommer före *z* i en sträng, så kommer *x* före *z*. Relationen ”kommer före” sägs därför vara (i) *irreflexiv*, (ii) *asymmetrisk* och (iii) *transitiv*. En sådan relation kallas ofta för en *sträng ordningsrelation*. Relationen ”kommer omedelbart före” är likaledes irreflexiv och asymmetrisk, men skiljer sig från ”kommer före” genom att vara intransitiv. (För en mera utförlig genomgång av begreppen reflexivitet, symmetri och transitivitet, se t. ex. Allwood, Andersson & Dahl (1977) *Logic in Linguistics*, Cambridge University Press.)

© Studentlitteratur

Kapitel 2 29 januari 1996 21.00:47 sida 20 av 26

(18) *min bror köpte cykeln*

(19) *av en gammal vän*

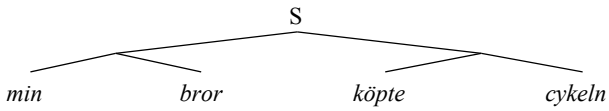
(20) *min bror köpte cykeln av en gammal vän*

2.2 Konstituentstruktur

De flesta lingvister är numera överens om att den syntaktiska strukturen hos en sats omfattar mer än bara den linjära strukturen hos en sträng av ord. De menar att orden i en sats (förutom att komma i en bestämd linjär ordning) är ordnade i större ordgrupper, eller *fraser*, som i sin tur bygger upp satser. Denna struktur kan åskådliggöras genom en *konstituentanalys*, som innebär att man delar upp satsen i dess *omedelbara konstituenter*, vilka i sin tur delas upp i sina *omedelbara konstituenter,* osv. ända ned till ordnivån.

Exempelvis består satsen *min bror köpte cykeln* av två omedelbara konstituenter, nämligen *min bror* och *köpte cykeln* (vilka ungefär motsvarar skolgrammatikens subjekt och predikat). Var och en av dessa konstituenter kan i sin tur delas upp i två delar, nämligen *min* och *bror* respektive *köpte* och *cykeln*. Därmed har vi kommit ned till ordnivån, som är konstituentanalysens slutpunkt. Analysen kan åskådliggöras med hjälp av ett s. k. *analysträd*, där vi använder symbolen S för att ange att det är en *sats* vi analyserar.

(21)



Konstituenter som består av flera ord kallas *fraser*, varav uttrycket *frasstruktur* som ofta används i stället för *konstituentstruktur*.

Än så länge kan det tyckas som om konstituentstrukturen (frasstrukturen) bara är ytterligare ett ”lager” med linjär struktur. Enligt detta synsätt skulle satsen *min bror köpte cykeln* kunna betraktas som en sträng av fonem (jfr 15), som en sträng av morfem (jfr 16), som en sträng av ord (jfr 17) eller som en sträng av fraser (jfr 21). Det finns emellertid en viktig skillnad mellan fraser å ena sidan och fonem, morfem och ord å den andra, och det är att fraser kan *innehålla andra fraser*. Detta innebär att vi inom en och samma

© Studentlitteratur

sats kan ha fraser på flera olika nivåer och att det i de flesta fall är omöjligt att analysera en sats som en enda sträng av fraser (medan det alltid finns en enda fonemsträng, en enda morfemsträng och en enda ordsträng). Man säger därför ofta att konstituentstrukturen är *hierarkisk*.

Antagandet att satser har en hierarkisk konstituentstruktur (utöver de linjära ord-, morfem- och fonemstrukturerna) grundar sig på två olika typer av observationer. För det första har man noterat att större ordgrupper i vissa satser tycks fylla funktioner som i andra satser fylls av enstaka ord.

(22) *han sov*

(23) *min bror sov*

(24) *min bror köpte cykeln*

(25) *den där gamla skurken köpte upp alla cyklar*

Om vi jämför satserna i (22–25) kan vi konstatera att *han* och *min bror* tycks ha samma roll i (22) och (23), medan *sov* och *köpte cykeln* i någon mening är ekvivalenta i (23) och (24).[4] På grundval av sådana observationer brukar man anta att *han* och *min bror* är fraser eller konstituenter av samma typ. Visserligen består *han* bara av ett enda ord, och det kan därför förefalla underligt att kalla det en fras, men i längden har det visat sig vara praktiskt att tillåta möjligheten att ett och samma element kan betraktas antingen som ett *ord* eller som en *fras*.[5] På samma sätt brukar man anta att *sov* och *köpte cykeln* är fraser av samma typ.

Om vi utsträcker analysen till (25), kan vi säga att *den där gamla skurken* är en fras av samma typ som *han* och *min bror* och att *köpte upp alla cyklar* är en fras av samma typ som *sov* och *köpte cykeln*, trots att de båda fraserna i (25) är mer komplexa och har omedelbara konstituenter som samtliga innehåller mer än ett ord (*den där gamla skurken* har de omedelbara konstituenterna *den där* och *gamla skurken*, medan *köpte upp alla cyklar* har de omedelbara konstituenterna *köpte upp* och *alla cyklar*). Tillsammans visar dessa exempel tydligt varför en sats konstituentstruktur generellt inte kan betrak-

[4] Ett ofta föreslaget kriterium på syntaktisk ekvivalens är möjlighet till *substitution* (utbyte) med bevarad grammatikalitet. Således kan vi i de aktuella exemplen byta ut *han* mot *min bror* (och vice versa) utan att satsernas grammatikalitet påverkas. På motsvarande sätt kan *sov* ersätta *köpte cykeln* (och vice versa) utan att resultatet blir ogrammatikaliskt.

[5] Detta är egentligen inte mera konstigt än att *ö* kan betraktas antingen som ett fonem eller som ett (enfonemigt) morfem; eller att *bil* kan betraktas antingen som ett morfem eller som ett (enmorfemigt) ord.

© Studentlitteratur

Kapitel 2 29 januari 1996 21.00:47 sida 22 av 26

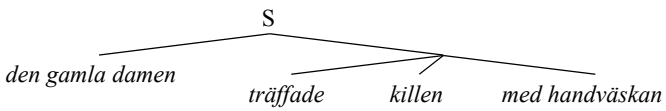
tas som en sträng av fraser. Vi kan visserligen betrakta (25) som en sträng bestående av fraserna *den där gamla skurken* och *köpte upp alla cyklar* (jfr 22), men vi kan lika gärna betrakta den som en sträng bestående av fraserna *den där*, *gamla skurken* och *köpte upp alla cyklar* (jfr 23), eller som en sträng bestående av fraserna *den där*, *gamla skurken*, *köpte upp* och *alla cyklar* (jfr 24). Dessutom måste vi komma ihåg att satser kan vara betydligt mer komplexa än (25) ovan, och att det i princip inte finns någon gräns för hur många frasnivåer en sats kan innehålla.

Den andra typen av observation som ligger bakom antagandet om en hierarkisk konstituentstruktur har att göra med *strukturell flertydighet*, dvs. att en och samma ordsträng kan representera olika satser, med olika frasstruktur och olika betydelse. Låt oss ta ett exempel.

(26) *den gamla damen träffade killen med handväskan*

Ordsträngen i (26) förefaller på ett systematiskt sätt vara förknippad med två olika tolkningar. I den första tolkningen fungerar frasen *med handväskan* som en bestämning till *träffade* och anger närmast det instrument med vars hjälp den gamla damen träffade killen (t. ex. i huvudet). I den andra tolkningen fungerar frasen *med handväskan* i stället som en bestämning till *killen* och anger ett attribut med vars hjälp vi kan identifiera den kille som den gamla damen träffade. Dessa båda tolkningar svarar enligt den gängse analysen mot *två olika satser*, som visserligen realiseras av *samma ordsträng* men som har *olika konstituentstruktur*. De två olika strukturerna illustreras i (27)[6] respektive (28).

(27)

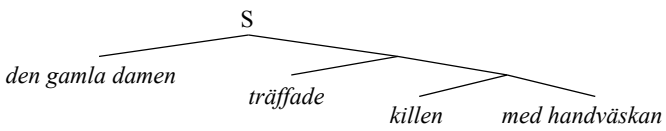


[6] (27) representerar en analys där de tre konstituenterna *träffade*, *killen* och *med handväskan* tillsammans utgör satsens predikatsdel (motsvarande *sov, köpte cykeln*, osv. i de tidigare exemplen). En alternativ analys är att *träffade* och *killen* tillsammans utgör frasen *träffade killen*, vilken i sin tur tillsammans med frasen *med handväskan* utgör satsens predikatsdel. Att konstruera ett analysträd som representerar den senare analysen lämnas som övning åt läsaren.

© Studentlitteratur

Kapitel 2 29 januari 1996 21.00:47 sida 23 av 26

(28)



Det är viktigt att notera att om vi verkligen vill säga att (27) och (28) representerar *olika* satser (och inte bara olika tolkningar av en och samma sats) så kan vi inte längre betrakta en sats enbart som en *grammatikalisk ordsträng*. I stället måste vi betrakta en sats som en *grammatikalisk ordsträng med en viss konstituentstruktur*. Detta är också det gängse satsbegreppet inom den generativa grammatiken.[7]

När vi tagit detta steg måste vi också konstatera att analysträd som (27) och (28) är en mera adekvat representationsform[8] för satser än den traditionella ”strängnotationen” i (26), eftersom den senare inte ger någon information om konstituentstruktur och därför bara utgör en partiell representation.

Innan vi lämnar denna avdelning vill vi göra ytterligare en anmärkning om begreppet *konstituentstruktur*. Vi sa tidigare att konstituentstrukturer skiljer sig från strängar genom att vara *hierarkiska*, dvs. konstituentstrukturer kan innehålla flera nivåer medan strängar alltid består av en enda strikt linjär följd av element. Av de exempel vi hittills gett kan man dock få intrycket att även om en konstituentstruktur som helhet inte är en sträng, så är varje konstituent som ingår i strukturen en sträng, dvs. i alla satser vi hittills har betraktat har alla konstituenter (på alla nivåer) varit sammanhängande delsträngar av satsens totala ordsträng. Det finns emellertid gott om satser som

---

[7] Man bör dock se upp med användningen av termen *sats* i litteraturen. Även om de flesta är överens om att det teoretiska begreppet sats bör inbegripa såväl ordsträng som konstituentstruktur, så används termen *sats* också i betydelsen *grammatikalisk ordsträng* (oavsett konstituentstruktur), t. ex. i formuleringar som ”satsen *den gamla damen träffade killen med handväskan* är strukturellt tvetydig”. Ibland används termen *sats* till och med i betydelsen ordsträng (oavsett grammatikalitet), t. ex. i formuleringar som ”satsen *han rakade mig själv* är ogrammatikalisk” (jfr fotnot 1).

[8] En alternativ notationsform är den s. k. parentes- eller klammernotationen, där konstituentstruktur indikeras med hjälp av parenteser (klamrar). De två aktuella satserna skulle t. ex. kunna representeras på följande sätt:

1. [[*den gamla damen*] [[*träffade*] [*killen*] [*med handväskan*]]]
2. [[*den gamla damen*] [[*träffade*] [[*killen*] [*med handväskan*]]]]

De två notationsformerna är i princip ekvivalenta, men trädnotationen är i allmänhet mera överskådlig (i synnerhet när strukturerna blir mer komplexa).

© Studentlitteratur

innehåller *diskontinuerliga konstituenter*, dvs. konstituenter som inte utgör sammanhängande strängar. Betrakta t. ex. följande exempel:

(29) *min bror köpte cykeln*

(30) *köpte min bror cykeln*

Enligt den analys vi presenterade i (21) ovan består (29) av de omedelbara konstituenterna *min bror* och *köpte cykeln*, vilka kan sägas utgöra satsens subjekts- respektive predikatsdel. Men vilka omedelbara konstituenter har vi i den motsvarande frågesatsen (30)? Den vanligaste analysen är att (30) innehåller exakt samma omedelbara konstituenter som (29), men att predikatsdelen i (30) är *diskontinuerlig*, dvs. *köpte* och *cykeln* utgör tillsammans predikatsdelen både i (29) och (30), men i (30) utgör de inte någon sammanhängande sträng.

Frågan om hur man exakt skall beskriva diskontinuerliga konstituenter har spelat en viktig roll i den generativa grammatikens utveckling och utgör alltjämt ett hett diskussionsämne. Tills vidare behöver vi dock bara notera att fenomenet existerar och att konstituentstrukturer (enligt det vanligaste synsättet) skiljer sig från strängar inte bara genom att vara hierarkiska utan också genom att i någon mening vara icke-linjära, dvs. det finns inget krav på att konstituenter måste utgöra sammanhängande strängar (även om det brukar anses vara normalfallet).

Vi kan därför säga att medan elementen i en sträng är ordnade av relationerna ”kommer före” och ”kommer omedelbart före” (jfr avdelning 2.1) så är elementen i en konstituentstruktur ordnade av (de icke-linjära) relationerna ”är konstituent i” och ”är omedelbar konstituent i”. Således gäller att satsen *min bror köpte cykeln* (jfr 21) har en konstituentstruktur där *min* och *bror* är konstituenter såväl i *min bror* som i *min bror köpte cykeln* men omedelbara konstituenter bara i den förra frasen. På samma sätt är *köpte* och *cykeln* konstituenter såväl i *köpte cykeln* som i *min bror köpte cykeln* men omedelbara konstituenter bara i den förra. Slutligen gäller att *min bror* och *köpte cykeln* är (omedelbara) konstituenter i *min bror köpte cykeln*.

2.3 Övningar

1. Försök analysera följande satser i deras omedelbara konstituenter. Fortsätt om möjligt analysen ända ned till ordnivå. Rita träd.

© Studentlitteratur

a. *Pelle älskar Lisa*
b. *Pelle älskar sin fru*
c. *katten stirrade på råttan*
d. *katten stirrade på råttan och råttan stirrade på katten*
e. *katten och råttan stirrade på varandra*
f. *jag gav honom en bok*
g. *jag gav min tjatiga kusin en bok om grammatik*
h. *det nya kaféet ligger på hörnet*
i. *på hörnet ligger det nya kaféet*
j. *lejonet åt upp Pelle*
k. *Pelle åts upp av lejonet*
l. *Pelle blev uppäten av lejonet*
m. *jag tror att Pelle vill ha en bok om grammatik*
n. *äntligen sover Pelle*
o. *äntligen har Pelle somnat*

2. Följande strängar av ord är tvetydiga. Diskutera om tvetydigheten kan förklaras som (fras)strukturell, och försök i så fall konstruera olika analysträd som motsvarar de olika betydelserna.

a. *Pelle kysste flickan i kiosken*
b. *Lisa lämnade Pelle olycklig*
c. *alla i det här rummet talar två språk*
d. *Pelle talade om hemligheten*
e. *Lisa jobbade på båten*

© Studentlitteratur

Kapitel 2 29 januari 1996 21.00:47 sida 26 av 26

3 Syntaktiska kategorier

En *struktur* kan betraktas som en mängd av *element* och ett antal *relationer* som råder mellan dessa element. I kapitel 2 tittade vi på två olika typer av struktur som man brukar tillskriva satser: *linjär struktur* och *konstituentstruktur*. I den linjära strukturen är elementen *ord*[1] och relationerna som råder mellan dem är ”kommer före” och ”kommer omedelbart före” (jfr avdelning 2.1). I konstituentstrukturen är elementen *ord* och *fraser* och relationerna som råder mellan dem är ”är konstituent i” och ”är omedelbar konstituent i”.

Hittills har vi alltså laborerat med språkliga element (ord, fraser) och relationer mellan dessa element. Så länge vi bara analyserar enstaka satser kan detta förefalla någorlunda tillfredsställande, men så snart vi vill formulera generella grammatiska regler som beskriver satserna i ett språk måste vi också ha möjlighet att tala om *kategorier* av språkliga element (dvs. kategorier av ord och fraser).[2] I detta kapitel skall vi därför utöka begreppsapparaten med syntaktiska *kategorier*.[3]

[1] Vi antar här att vi i första hand är intresserade av den *syntaktiska* strukturen (och lämnar följaktligen den morfologiska och fonologiska strukturen därhän).
[2] Indirekt uppkom detta behov redan i kapitel 2 då vi jämförde olika fraser (t. ex. *han*, *min bror* och *den där gamla skurken*) och hävdade att de var av samma *typ* (dvs. av samma kategori).
[3] Eftersom de minsta enheter vi kategoriserar är *ord* är beteckningen *syntaktiska kategorier* bättre än *grammatiska kategorier* som också bör kunna omfatta t. ex. morfemkategorier.

© Studentlitteratur

3.1 Intrinsikala och relationella kategorier

Den traditionella skolgrammatiken innehåller två olika (men besläktade) klassifikationssystem för språkliga uttryck: *ordklasser* och *satsdelar*. Båda dessa system kan användas för att klassificera satsers beståndsdelar.

(31) *hunden gömde benet*

Om vi betraktar (31) som en sammanhängande sats kan vi för det första ta ut satsdelar och säga att *hunden* är subjekt, *gömde* predikat(sverb) och *benet* (direkt) objekt. För det andra kan vi klassificera de ingående orden genom att säga att *hunden* och *benet* är substantiv medan *gömde* är ett verb.

Om vi däremot betraktar orden *hunden*, *gömde* och *benet* oberoende av deras förekomst i en bestämd sats, kan vi bara använda den senare kategoriseringen (ordklasskategoriseringen).

Detta beror på att ordklasser (enligt den mest traditionella uppfattningen) är *intrinsikala kategorier*, dvs. ett ords ordklasstillhörighet tillhör ordet självt och är oberoende av dess förekomst i någon specifik syntaktisk struktur. Satsdelar är i stället *relationella kategorier*, dvs. de betecknar en viss relation som ord (eller fraser) kan ha till andra ord (fraser) i en bestämd syntaktisk struktur.[4]

Inom den generativa grammatiken har man i allmänhet utgått från *intrinsikala kategorier* och försökt beskriva hur dessa kan kombineras till syntaktiska strukturer. Relationella kategorier har antingen inte behandlats alls eller definierats i termer av intrinsikala kategorier och strukturella konfigurationer.[5] Ofta använder man därför termen *syntaktisk kategori* enbart för att beteckna intrinsikala syntaktiska kategorier, medan de relationella kategorierna benämns *syntaktiska funktioner*. Vi kommer inte att ha mycket att säga om syntaktiska funktioner i denna bok.

[4] Observera dock att ett ords ordklasstillhörighet indirekt är förknippad med en viss potential att fylla olika satsdelsfunktioner (t. ex. substantiv – subjekt, objekt; (transitivt) verb – predikat; adjektiv – attribut, predikatsfyllnad; adverb – adverbial).

[5] Det finns dock teorier som avviker från detta mönster genom att anta de relationella kategorierna som primitiva. De viktigaste torde vara *Relational Grammar*, *Lexical-Functional Grammar* och (eventuellt) *Head-Driven Phrase Structure Grammar*.

© Studentlitteratur

Kapitel 3 29 januari 1996 21.05:23 sida 28 av 32

3.2 Lexikala och icke-lexikala kategorier

Bland de intrinsikala syntaktiska kategorierna kan vi urskilja två viktiga huvudtyper: *lexikala kategorier* (eller *ordklasser*) och *icke-lexikala kategorier* (eller *frasklasser*). För att illustrera denna distinktion kan vi steg för steg analysera strukturen hos följande välbekanta sats.

(32) *min bror köpte cykeln av en gammal vän*

Till att börja med kan (32) analyseras som en sträng av ord, vilka vart och ett kan tilldelas en bestämd ordklass enligt följande:

(33)

DET	N	TV	N	P	DET	A	N
min	*bror*	*köpte*	*cykeln*	*av*	*en*	*gammal*	*vän*

De kategorisymboler som förekommer i (34) läses ut på följande sätt: DET = *determinator* (eng. *determiner*), en ordklass som omfattar artiklar och vissa förenade pronomen och som saknar direkt motsvarighet bland de traditionella ordklasserna; N = *nomen* (eng. *noun*) eller *substantiv*; TV = *transitivt verb*, dvs. en undergrupp av den traditionella ordklassen *verb*; P = *preposition*; och A = *adjektiv*. Alla dessa kategorier är alltså exempel på *lexikala kategorier*, dvs. kategorier vars medlemmar är *lexikala enheter* (eller *ord*).

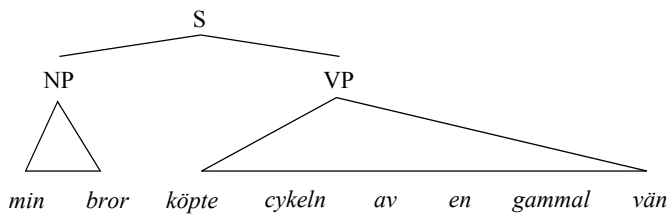
Om vi så går vidare till satsens konstituentstruktur, kan vi först urskilja de två omedelbara konstituenterna *min bror* och *köpte cykeln av en gammal vän*. Den förra är uppbyggd kring ett substantiv (eller nomen), nämligen *bror*, den senare kring ett (transitivt) verb, nämligen *köpte*. Fraser av dessa typer kallas vanligen *nominalfras* respektive *verbfras*, vilket förkortas NP respektive VP (efter engelskans *noun phrase* respektive *verb phrase*). Om vi sätter in dessa etiketter i ett analysträd får vi följande (partiella) analys av satsens konstituentstruktur:[6]

[6] Trianglarna i (34) markerar att konstituentanalysen är ofullständig vad beträffar nominalfrasens och verbfrasens inre struktur.

© Studentlitteratur

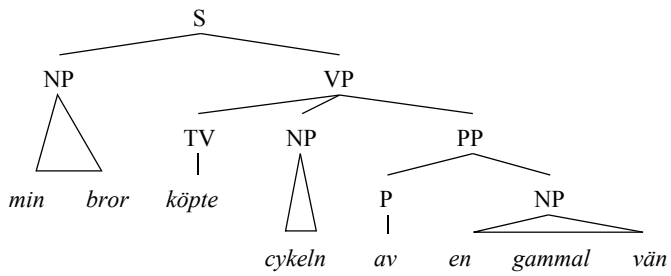
Kapitel 3 29 januari 1996 21.05:23 sida 29 av 32

(34)



Verbfrasen *köpte cykeln av en gammal vän* har tre omedelbara konstituenter: det transitiva verbet *köpte*, nominalfrasen *cykeln* samt frasen *av en gammal vän*, som är en *prepositionsfras* (PP). Denna prepositionsfras har i sin tur två omedelbara konstituenter: prepositionen *av* och nominalfrasen *en gammal vän*. Detta ger oss följande:

(35)



Kvar att analysera är de tre nominalfraserna *min bror*, *cykeln* och *en gammal vän*. Låt oss först notera att *cykeln* i denna struktur måste betraktas dels som ett *ord* (ett substantiv), dels som en *fras* (en nominalfras). Anledningen är att *cykeln* är syntaktiskt ekvivalent med (dvs. kan fylla samma funktioner som) mer komplexa nominalfraser som t. ex. *den gamla rostiga cykeln* och *cykeln som du lånade* (jfr *min bror köpte den gamla rostiga cykeln av en gammal vän* respektive *min bror köpte cykeln som du lånade av en gammal vän*). Som omedelbar konstituent i verbfrasen *köpte cykeln av en gammal vän* är *cykeln* därför en nominalfras, och denna nominalfras har i sin tur en enda omedelbar konstituent, nämligen substantivet *cykeln*.

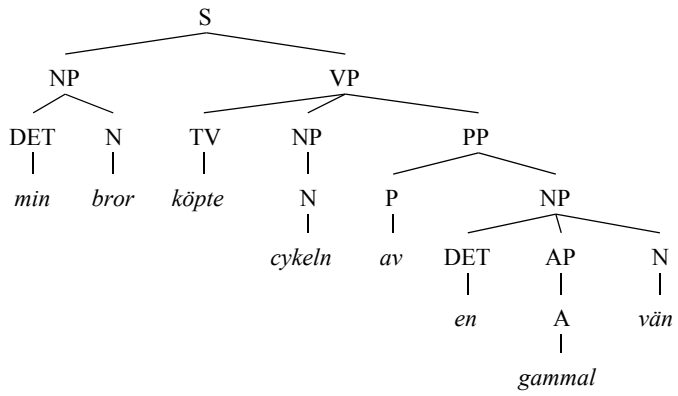
Om vi går vidare till de båda andra nominalfraserna, kan vi tills vidare anta att *min bror* har de två omedelbara konstituenterna *min* och *bror*, medan *en gammal vän* har de tre omedelbara konstituenterna *en*, *gammal* och *vän*. Vidare förefaller det klart att *min* och *en* bör kategoriseras som determinatorer, och *bror* och *vän* som substantiv. Däremot måste konstituenten *gammal*

© Studentlitteratur

Kapitel 3 29 januari 1996 21.05:23 sida 30 av 32

behandlas parallellt med *cykeln* ovan, dvs. den måste betraktas både som ett ord och som en fras, närmare bestämt som en adjektivfras, eftersom den är syntaktiskt ekvivalent med mer komplexa fraser som *mycket gammal* och *alldeles förskräckligt gammal* (jfr *en mycket gammal vän*, *en alldeles förskräckligt gammal vän*). Detta ger oss till sist följande fullständiga analysträd för satsen *min bror köpte cykeln av en gammal vän*, med alla syntaktiska kategorier inplacerade.

(36)



Sammanfattningsvis har vi alltså urskiljt två typer av kategorier: *lexikala kategorier* (ordklasser), t. ex. N, TV, A, P, DET; och *icke-lexikala kategorier* (frasklasser), t. ex. S, NP, VP, AP, PP. Observera att fraser av kategorierna NP, VP, AP och PP (men inte S) typiskt innehåller ett ord av kategorin N, V, A respektive P. Denna (obligatoriska) konstituent kallas ofta frasens *huvudkonstituent* (ibland *huvudord* eller kort och gott *huvud*). Generellt kan vi därför säga att en fras av kategori XP har som huvud ett ord av kategori X (där X = N, V, A eller P).[7] När en fras består av ett enda ord (dvs. bara har en enda konstituent) är detta ord alltid huvudordet (jfr (36) där *cykeln* (N) är huvud och enda konstituent i en nominalfras, medan *gammal* (A) är huvud och enda konstituent i en adjektivfras).

[7] Utöver ovannämnda fraskategorier förekommer ibland också ADVP (adverbfras) med ett ord av kategorin ADV (adverb) som huvud.

© Studentlitteratur

Kapitel 3 29 januari 1996 21.05:23 sida 31 av 32

3.3 Övningar

1. Försök sätta ut kategorisymboler i de analysträd du konstruerade i samband med övning 1 och 2 i kapitel 2. Du behöver förmodligen hitta på en del egna kategorisymboler för fras- och ordklasser som vi hittills inte behandlat. (Dessa kategorier kommenteras i lösningsförslagen till övningarna.)

© Studentlitteratur

Kapitel 3 29 januari 1996 21.05:23 sida 32 av 32

4 Analysträd

I kapitel 2 och 3 har vi steg för steg utvecklat ett satsbegrepp som kan beskrivas på följande sätt. En sats kan betraktas som en sträng av ord, där varje ord tillhör en bestämd ordklass (lexikalisk kategori). Orden i en sådan sträng är ordnade av relationerna ”kommer före” och ”kommer omedelbart före”. En sats har dessutom en konstituentstruktur, dvs. den är uppbyggd av omedelbara konstituenter (fraser), vilka i sin tur är uppbyggda av omedelbara konstituenter osv. ända ned till ordnivån. Elementen i denna konstituentstruktur är relaterade genom relationerna ”är konstituent i” och ”är omedelbar konstituent i”. Varje fras tillhör en bestämd frasklass (icke-lexikal kategori) och de flesta fraser har ett huvud som tillhör en bestämd ordklass (N är t. ex. huvud i NP, A i AP osv.).

Vi har också sett hur satser kan representeras med hjälp av *analysträd*. I detta kapitel tänkte vi göra en liten utvikning och diskutera i mera formella termer dels vad för slags objekt analysträd är och dels på vilket sätt de kan sägas representera syntaktiska strukturer.

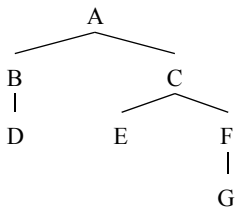
4.1 Noder och grenar

Formellt sett består ett analysträd av ett antal *noder* (lat. ”knutar”) som förbinds av *grenar*. Om det går en gren från en nod A till en annan nod B, så säger vi att A *omedelbart dominerar* B. Observera att relationen ”dominerar omedelbart” är irreflexiv och asymmetrisk, dvs. ingen nod dominerar omedelbart sig själv och om A omedelbart dominerar B så gäller att B *inte* omedelbart dominerar A. Detta skulle man kunna markera genom att utforma grenarna som enkelriktade pilar, men det vanligaste är att man låter det framgå på ett mera indirekt sätt genom att dominerande noder placeras högre i trädet än dominerade. I (37) är det alltså så att nod A omedelbart dominerar nod B och nod C, att nod B omedelbart dominerar nod D, att nod

© Studentlitteratur

C omedelbart dominerar nod E och nod F, samt att nod F omedelbart dominerar nod G.

(37)



De träd vi kommer att intressera oss för uppfyller dessutom följande två krav: de har en unik *rot*, vilket innebär att det finns exakt en nod som inte omedelbart domineras av någon annan nod (i 37 finns exakt en sådan nod, nämligen A); varje nod (förutom roten) har exakt en *moder*,dvs. domineras omedelbart av exakt en annan nod (detta krav är också uppfyllt i (37), där A är moder till B och C, B till D, C till E och F, och F till G).

Lite terminologi: Om A omedelbart dominerar B, så sägs A vara *moder* till B och B *dotter* (eller *barn*) till A. Om A och B har samma moder, så är de *systrar* (eller *syskon*). Noder som saknar döttrar kallas *terminala* noder.

Vid sidan av relationen ”dominerar omedelbart” finns också den transitiva relationen ”dominerar”. I (37) gäller t. ex. att nod A dominerar samtliga övriga noder, att B dominerar D, att C dominerar E, F och G, samt att F dominerar G. Av tekniska skäl (som vi inte behöver gå in på här) brukar man dessutom låta relationen ”dominerar” (men inte ”dominerar omedelbart”) vara reflexiv, dvs. man antar att varje nod dominerar sig själv.

När trädstrukturer av ovanstående typ används för att representera språkliga uttryck (i synnerhet satser) är tanken att *roten* skall representera hela uttrycket (satsen), att varje övrig *nod* skall representera en *konstituent* i det språkliga uttrycket (på någon nivå), att *terminala noder* skall representera de minsta konstituenterna (dvs. vanligtvis de enskilda orden), och slutligen att relationerna ”dominerar” och ”dominerar omedelbart” (som representeras med hjälp av grenar i trädet) skall motsvara relationerna ”är konstituent i” och ”är omedelbar konstituent i” (jfr avdelning 2.2).

Ett trädstruktur som (37) skulle alltså kunna representera en sats A med två omedelbara konstituenter B och C, där B består av ett enda ord D, medan C består av ett ord E och en fras F, och där F slutligen består av ett enda ord G. En sådan sats är t. ex. *han köpte cykeln* (A = *han köpte cykeln*, B = *han*

© Studentlitteratur

(fras), C = *köpte cykeln*, D = *han* (ord), E = *köpte*, F = *cykeln* (fras), G = *cykeln* (ord)).

4.2 Etiketterade noder

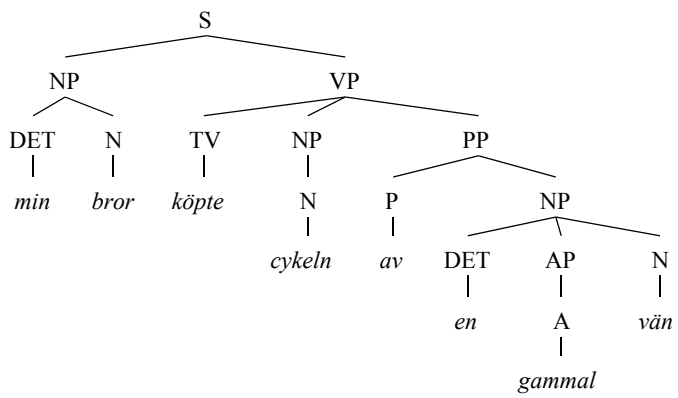
Typiskt för de analysträd man använder inom den generativa grammatiken är emellertid också att noderna är ”etiketterade”, dvs. att varje nod i trädet är märkt med en symbol av något slag. Dessa symboler kan indelas i två huvudkategorier: *terminala symboler*, som representerar enskilda ord; och *icke-terminala symboler*, som betecknar syntaktiska kategorier.

De *terminala* symbolerna kallas så därför att de förekommer längst ut på trädens grenar, dvs. de fungerar som etiketter på noder som inte har några döttrar; de *icke-terminala* symbolerna kallas så därför att de *inte* förekommer längst ut på grenarna.

I (38) förekommer de terminala symbolerna *min*, *bror*, *köpte*, *cykeln*, *av*, *en*, *gammal* och *vän*, samt de icke-terminala symbolerna S, NP, VP, PP, AP, DET, N, TV, P och A. Det är viktigt att notera att en och samma symbol kan fungera som etikett på flera olika noder. I (38) förekommer t. ex. symbolen NP på tre olika ställen, dvs. det finns tre olika noder som har samma etikett (vilket motsvarar det faktum att den sats som representeras av (38) innehåller tre fraser av samma typ, nämligen *min bror*, *cykeln* och *en gammal vän*). Det är därför viktigt att hålla isär å ena sidan *noderna* och å andra sidan *nodernas etiketter*; de förra representerar *konstituenter*, de senare *kategorier som konstituenterna tillhör*. Andra exempel på symboler som förekommer flera gånger i (38) är N (tre gånger) och DET (två gånger).

© Studentlitteratur

(38)



De *icke-terminala* symbolerna kan indelas vidare i *preterminala*, dvs. de som förekommer näst längst ut på grenarna, närmast före de terminala symbolerna, och *övriga*. Denna distinktion är tänkt att motsvara den ovan gjorda distinktionen mellan *lexikala kategorier* (ordklasser) och *icke-lexikala kategorier* (frasklasser). I (38) är således DET, N, TV, P och A preterminala (lexikala) symboler, medan S, NP, VP, PP och AP är övriga (icke-lexikala) symboler.

4.3 Precedensrelationer i analysträd

Hittills i detta kapitel har vi tittat på hur analysträd representerar satsers *konstituentstruktur* genom dominansrelationer (grenar) mellan konstituenter (noder) av bestämda kategorier (med bestämda etiketter).

Lingvister har dock i allmänhet även velat representera *linjär struktur* i sina analysträd. Detta åstadkoms genom att man antar att de *terminala* noderna i ett träd är ordnade av relationen ”kommer före” i den ordning de skrivs från vänster till höger, dvs. man antar att de terminala noderna representerar en sträng vars element är de ord som representeras av de terminala symbolerna. (I 38 ovan representerar de terminala noderna således strängen *han köpte cykeln av en gammal vän*.)

Relationen ”kommer före” kallas ibland (i samband med analysträd) för *linjär precedens* (eng. *linear precedence*), och relationerna ”kommer före”

© Studentlitteratur

Kapitel 4 29 januari 1996 21.07:02 sida 36 av 38

och ”kommer omedelbart före” kallas ofta med en gemensam term för *precedensrelationer*.

För att fånga den linjära strukturen hos satser skulle det i princip räcka med att låta de terminala noderna vara ordnade av precedensrelationer medan övriga noder kunde förbli (linjärt) oordnade. Traditionellt har man emellertid utsträckt precedensrelationerna till att även gälla mellan (vissa) icke-terminala noder enligt följande principer:

(39) En nod A *kommer före* en nod B om och endast om den sekvens av terminalnoder som domineras av A *kommer före* den sekvens av terminalnoder som domineras av B och de båda sekvenserna ej är överlappande.

(40) En nod A *kommer omedelbart före* en nod B om och endast om den sekvens av terminalnoder som domineras av A *kommer omedelbart före* den sekvens av terminalnoder som domineras av B (utan överlapp).

För att dessa definitioner skall ge önskat resultat måste vi dessutom lägga till antagandet att relationen dominans (i motsats till omedelbar dominans) är reflexiv, dvs. att varje nod dominerar sig själv (jfr avdelning 4.1).

Om vi återvänder till (38) kan vi nu t. ex. konstatera att den terminala noden längst till vänster (med den terminala symbolen *min*) *kommer före* samtliga övriga noder utom tre: modern (märkt DET), ”mormodern” (märkt NP), samt roten (märkt S). Vidare gäller t. ex. att den preterminala noden märkt P *kommer före* de terminala noder som är märkta *en*, *gammal* och *vän*, dessas mödrar (märkta DET, A respektive N), samt A-nodens moder och mormoder (märkta AP respektive NP).

En viktig konsekvens av att relationen *kommer före* definieras på detta sätt är att varje par av noder X, Y i ett analysträd måste vara relaterade på exakt ett av följande fyra sätt: X dominerar Y, Y dominerar X, X kommer före Y eller Y kommer före X.

© Studentlitteratur

4.4 Övningar

1. Gå igenom (38) och avgör för varje par av noder X, Y på vilket sätt de två noderna är relaterade (dvs. om X dominerar Y, Y dominerar X, X kommer före Y eller Y kommer före X). Håll på så länge du orkar eller tills du tycker att du fattat galoppen. (Det finns sammanlagt 253 par av noder.)

2. Återvänd till de analysträd du konstruerade i samband med övning 1 i kapitel 3 ovan och försök för varje träd indela noderna i följande kategorier:

   a. *terminala*
   b. *icke-terminala*
   c. *preterminala*
   d. *icke-terminala men ej preterminala*

3. Utgå från den klassificering du gjort i övning 2 och försök klassificera även de symboler du använt för att etikettera noderna i dina träd. Använd följande kategorier:

   a. *terminala symboler (lexikonenheter, ord)*
   b. *lexikala kategorisymboler (ordklasser)*
   c. *icke-lexikala kategorisymboler (frasklasser)*

© Studentlitteratur

Kapitel 4 29 januari 1996 21.07:02 sida 38 av 38

5 Frasstrukturgrammatik

I kapitel 1 slog vi fast att en generativ grammatisk beskrivning av ett enskilt språk är en explicit redogörelse för mängden av satser i språket. Denna redogörelse måste åtminstone definiera mängden av satser (dvs. generera alla och endast de grammatikaliska satserna) och beskriva satsernas struktur.

I kapitel 2 konstaterade vi att satser har åtminstone två olika sorters (syntaktisk) struktur: linjär struktur och konstituentstruktur, och i kapitel 3 och 4 har vi sedan successivt utvecklat en begreppsapparat för att representera dessa strukturer. Speciellt har vi försökt visa att etiketterade analysträd, som innehåller både dominans- och precedensrelationer, kan användas för att representera satser på ett sätt som fångar både den linjära ordstrukturen och den hierarkiska konstituentstrukturen.

Ett sätt att åstadkomma en adekvat grammatisk beskrivning av ett språk skulle därför vara att försöka definiera mängden av välformade analysträd för språket i fråga, dvs. att försöka definiera en mängd som innehåller exakt ett analysträd för varje grammatikalisk sats i språket. Detta är också den metod som varit gängse inom den generativa grammatiken.[1]

I detta kapitel skall vi titta på en klassisk modell för grammatisk beskrivning, som kan användas just för att definiera analysträd. Modellen, som går under beteckningen frasstrukturgrammatik,[2] är relativt enkel och spelar (i sin traditionella form) ingen framträdande roll i de moderna generativa teorierna, men den ligger till grund för de senaste fyrtio årens utveckling, och

[1] På senare år har analysträd fått konkurrens av en annan sorts representationer som brukar kallas *särdragsstrukturer* (se kapitel 8). Fortfarande spelar dock analysträd en central roll inom de flesta grammatiska teorier.

[2] I mera tekniska sammanhang benämner man denna typ av grammatik *kontextfri* (frasstruktur)grammatik, för att skilja den från andra, liknande typer av grammatiker. Dessa teknikaliteter kan vi dock bortse från i detta sammanhang.

© Studentlitteratur

den tillhör utan tvekan det man kan kalla elementa inom generativ grammatik.

5.1 Frasstrukturregler

En frasstrukturgrammatik för ett enskilt språk består i princip av två delar: ett *lexikon,* dvs. en lista över orden i språket där varje ord tilldelas en syntaktisk kategori, och en uppsättning *frasstrukturregler,* dvs. regler som definierar analysträd vars noder är etiketterade med syntaktiska kategorisymboler. Låt oss börja med att titta på frasstrukturreglerna.

Frasstrukturregler består av två delar: ett *vänsterled* och ett *högerled.* Vänsterledet är en syntaktisk kategorisymbol, och högerledet är en sekvens av syntaktiska kategorisymboler. Den vanliga notationen för frasstrukturregler är att man skriver vänsterledet, följt av en högerriktad pil, följt av högerledet, på det sätt som illustreras i (41–43).

(41) S -> NP VP

(42) NP -> N

(43) VP -> TV NP

Det finns flera olika sätt att tolka frasstrukturregler, men vi skall till att börja med nöja oss med att ta upp två olika tolkningar, som haft betydelse för den generativa grammatikens utveckling. Låt oss kalla dem *strängderivationstolkningen* respektive *trädkonstruktionstolkningen.*

5.2 Strängderivation

Strängderivationstolkningen är den äldsta och på sätt och vis den som är minst intressant ur lingvistisk synvinkel.[3] Den går ut på att frasstrukturreg-

[3] Den frasstrukturgrammatiska modellen har inte bara användningar inom lingvistik utan också inom t. ex. datalogi (i beskrivningen av programmeringsspråk). Det är naturligtvis inte säkert att den syn på frasstrukturregler som är mest intressant för lingvistikens vidkommande också är mest intressant ur andra perspektiv

© Studentlitteratur

Kapitel 5 29 januari 1996 21.36:05 sida 40 av 48

ler tolkas som regler för att *härleda* (eller *derivera*) strängar på ungefär samma sätt som i ett logiskt-matematiskt bevis.[4] Ur detta perspektiv kan reglerna i (41–43) tolkas på följande sätt:

(44) Om du har en sträng som innehåller symbolen S, så får du lov att härleda den sträng som är exakt som den ursprungliga strängen förutom att symbolen S är utbytt mot en sträng av symboler vars första element är symbolen NP och vars andra element är symbolen VP.

(45) Om du har en sträng som innehåller symbolen NP, så får du lov att härleda den sträng som är exakt som den ursprungliga strängen förutom att symbolen NP är utbytt mot symbolen N (egentligen: den sträng av symboler vars enda element är symbolen N).

(46) Om du har en sträng som innehåller symbolen VP så får du lov att härleda den sträng som är exakt som den ursprungliga strängen förutom att symbolen VP är utbytt mot en sträng av symboler vars första element är symbolen TV och vars andra element är symbolen NP.

Om vi tänker oss att vi utgår från symbolen S (eller rättare sagt från den sträng vars enda element är symbolen S), så kan vi genom att tillämpa reglerna i (41–43) åstadkomma följande härledning:

(47)

S	
NP VP	(regel 41)
N VP	(regel 42)
N TV NP	(regel 43)
N TV N	(regel 42)

Enligt strängderivationstolkningen definierar en frasstrukturgrammatik (dvs. en uppsättning frasstrukturregler) en mängd av *strängar*, nämligen exakt de strängar som kan härledas enligt grammatikens regler givet en viss *startsymbol* (dvs. en symbol att börja härledningarna med).[5]

I en frasstrukturgrammatik som innehåller reglerna i (41–43) och som har S som startsymbol kan vi alltså bl. a. härleda strängen N TV N. Om vi tolkar detta som en grammatisk beskrivning, blir innebörden att en sträng bestå-

[4] För en orientering om logiska härledningssystem (deduktionssystem), se t. ex. Allwood, Andersson & Dahl (1977) *Logic in Linguistics*, Cambridge University Press.
[5] I mera formella sammanhang kallas startsymbolen ofta *axiom*.

© Studentlitteratur

Kapitel 5 29 januari 1996 21.36:05 sida 41 av 48

ende av ett substantiv (N), följt av ett transitivt verb (TV), följt av ett substantiv (N) är en grammatikalisk sats (S) i det språk som beskrivs. Om språket vi beskrev t. ex. var svenska, skulle härledningen i (47) vara relevant för en sats som *brorsan köpte cykeln.*

Lingvister har emellertid inte visat något större intresse för denna tolkning av frasstrukturgrammatikens regler, eftersom *strängar* allmänt anses vara otillräckliga för att representera satsers struktur. Låt oss därför lämna denna tolkning och i stället betrakta vad vi kallat för *trädkonstruktionstolkningen.*

5.3 Trädkonstruktion

I stället för att tolka reglerna i (41–43) som regler för hur man härleder strängar, kan vi tolka dem som regler för hur man konstruerar analysträd. Med detta synsätt får reglerna i (41–43) följande innebörd:

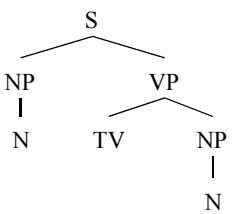
(48) Om du har en nod med etiketten S, så får du lov att lägga till två döttrar (dvs. du får lov att dra grenar från S-noden till två nya noder), som har etiketterna NP respektive VP, och där NP-noden kommer före VP-noden.

(49) Om du har en nod med etiketten NP, så får du lov att lägga till en dotter, som har etiketten N.

(50) Om du har en nod med etiketten VP, så får du lov att lägga till två döttrar, som har etiketterna TV respektive NP, och där TV-noden kommer före NP-noden.

Om vi tänker oss att vi börjar med en startnod (i stället för en startsträng, jfr avdelning 5.2), som har etiketten S, kan vi genom att tillämpa reglerna i (41–43) konstruera följande träd:

(51)



© Studentlitteratur

Kapitel 5 29 januari 1996 21.36:05 sida 42 av 48

Enligt trädkonstruktionstolkningen definierar en frasstrukturgrammatik en mängd av *etiketterade analysträd,* nämligen exakt de analysträd som kan konstrueras enligt grammatikens regler, givet en nod etiketterad med grammatikens startsymbol.

I en frasstrukturgrammatik som innehåller reglerna i (41–43) och som har S som startsymbol kan vi alltså bl. a. konstruera (51). Om vi tolkar detta som en grammatisk beskrivning, blir innebörden att en grammatikalisk sats (S) kan bestå av en nominalfras (NP) följd av en verbfras (VP), att en verbfras kan bestå av ett transitivt verb (TV) följt av en nominalfras, och att en nominalfras (vare sig den är omedelbar konstituent i en sats eller i en verbfras) kan bestå av ett substantiv (N). Återigen kan vi på svenska jämföra med en sats som *brorsan köpte cykeln.*

Vi hoppas det är klart vid det här laget varför trädkonstruktionstolkningen är intressantare för lingvister än strängderivationstolkningen. Med denna tolkning ger oss frasstrukturgrammatiken en metod för att definiera analysträd, vilket var precis vad vi önskade i inledningen till detta kapitel. Vad som saknas än så länge är emellertid ett sätt att också få in terminala symboler i träden, dvs. symboler som representerar de enskilda orden i språket. För detta behöver vi ett *lexikon* och en *lexikal insättningsregel.*

5.4 Lexikon

Ett *lexikon* i en klassisk frasstrukturgrammatik är en lista över de minsta uttrycken i det språk som skall beskrivas, dvs. de uttryck som ej antas vara sammansatta av mindre uttryck. Dessa uttryck kan vi med en generell term kalla *lexikonenheter* eller *lexikala uttryck.* Om vi alltjämt utgår från *ordet* som den minsta beståndsdelen i satsen (dvs. om vi fortsätter att intressera oss enbart för satsens syntaktiska struktur), så blir *lexikala uttryck* detsamma som *ord* och ett lexikon bör alltså vara en lista över alla *ord* i det språk vi skall beskriva. Dessutom måste varje lexikonenhet tilldelas en *syntaktisk kategori* (närmare bestämt en *lexikal kategori*; jfr kapitel 3).[6]

Lexikonet i en frasstrukturgrammatik för svenska skulle således bl. a. kunna innehålla följande enheter:[7]

© Studentlitteratur

(52) DET: min
DET: en
DET: den
N: bror
N: cykeln
N: vän
N: damen
N: killen
N: handväskan
PRO: han
A: gammal
A: gamla
IV: sov
TV: köpte
TV: träffade
P: av
P: med

För att koppla ihop lexikonet med frasstrukturreglerna behöver vi dessutom en s. k. *lexikal insättningsregel,* dvs. en regel som talar om hur vi får sätta in lexikonenheter som etiketter på noder i analysträd. Denna kan t. ex. formuleras på följande sätt:

(53) Om *a* är en lexikonenhet och *x* är en nod utan döttrar i ett analysträd, och om etiketten på *x* betecknar den lexikala kategori som *a* tillhör, så får du lov att lägga till en dotter till *x* med etiketten *a*.[8]

Vad (53) säger är att vi får sätta in en lexikonenhet (ett ord) som etikett på dottern till en (preterminal) nod om deras kategorier överensstämmer, dvs. om lexikonenheten tillhör den syntaktiska kategori som betecknas av nodens etikett, och om noden inte redan har en dotter. (Det senare villkoret

[6] Information om syntaktisk kategori för varje lexikonenhet kan sägas vara det minimum av information ett lexikon måste ge, och i den klassiska frasstrukturgrammatiken är det också den enda information som lexikonet ger. Detta kan jämföras med de böcker som vi i dagligt tal kallar ”lexikon”, vilka normalt innehåller betydligt mer information, t. ex. om fonologi (uttal), morfologi (böjning) och semantik (betydelse). Hur mycket av denna information som också bör inkorporeras i lexikonet för en generativ grammatik, och på vilket sätt det skall göras, är ännu en öppen fråga, men i den klassiska frasstrukturgrammatiken nöjer vi oss alltså med information om syntaktisk kategori.

[7] Observera att olika former av samma ord (t. ex. *gammal* och *gamla*) måste räknas upp som skilda lexikonenheter så länge vi inte tar hänsyn till morfologi.

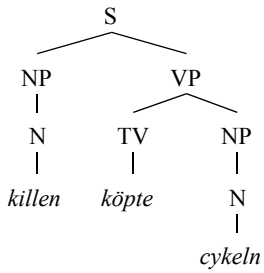
© Studentlitteratur

Kapitel 5 29 januari 1996 21.36:05 sida 44 av 48

behövs för att utesluta möjligheten att sätta in flera lexikonenheter under samma preterminala nod.)

I förhållande till lexikonet i (52) säger alltså (53) att vi får sätta in *min, en* eller *den* som etikett på dottern till en DET-nod, att vi får sätta in *bror, cykeln, vän, damen, killen* eller *handväskan* som etikett på dottern till en N-nod osv. Om vi utgår från trädet i (51), är t. ex. (54) en tillåten lexikal insättning (och följaktligen ett välformat analysträd givet 41–43, 52 och 53).

(54)



En viktig egenskap hos ett träd som (54) är att alla *terminala noder* (dvs. noder utan döttrar) har etiketter som är *lexikala uttryck* (eller *terminala symboler*). Ett sådant träd sägs vara *terminerat* (eng. *terminated)* eller *avslutat.* Klassen av terminerade analysträd är en viktig klass, eftersom det endast är dessa som kan representera enskilda satser. Ett exempel på ett icke-terminerat analysträd är (51), som därför inte heller representerar någon enskild sats.[9]

[8] Vi har har valt att formulera den lexikala insättningsregeln direkt i ”trädkonstruktiva” termer. En mera neutral formulering är att uttrycka den som ett regelschema: För varje lexikonenhet *a* med lexikal kategori *A*, så finns det en frasstrukturregel *A -> a*. Ett sådant regelschema skulle alltså givet ett visst lexikon ”generera” en frasstrukturregel för varje lexikonenhet (i förhållande till (52) skulle den t. ex. generera DET -> *min* och N -> *bror).* Dessa ”genererade” regler skulle sedan kunna tolkas på samma sätt som de övriga frasstrukturreglerna (antingen som strängderivationsregler eller som trädkonstruktionsregler).

[9] Däremot kan man möjligen säga att ett träd som (51) representerar en mängd av satser, nämligen de satser som har den syntaktiska struktur som representeras av (51) och som kan representeras av olika terminerade varianter av (51), såsom t. ex. (54).

© Studentlitteratur

Kapitel 5 29 januari 1996 21.36:05 sida 45 av 48

5.5 En enkel frasstrukturgrammatisk modell

Efter att ha tittat på de huvudsakliga komponenterna i en frasstrukturgrammatik tänkte vi nu försöka formulera en mer exakt modell för grammatisk beskrivning. I nästa kapitel tänkte vi sedan exemplifiera modellen genom att formulera en beskrivning av ett mycket litet fragment av svenska.

En *frasstrukturgrammatik* (enligt vår modell) består av två komponenter: ett *lexikon* och en uppsättning *frasstrukturregler.* Ett *lexikon* är en lista av ordnade par, där varje par består av en *lexikal kategorisymbol* och ett *ord.* Till lexikonet räknar vi också en *lexikal insättningsregel* av den typ vi formulerade i (53) ovan.

Förrådet av *syntaktiska kategorisymboler* indelas i två disjunkta[10] delmängder: *lexikala* och *icke-lexikala.* De lexikala kategorisymbolerna används dels i lexikonet (för att specificera lexikonenheternas syntaktiska kategoritillhörighet, se ovan), dels i frasstrukturreglerna (se nedan). De icke-lexikala kategorisymbolerna används enbart i frasstrukturreglerna. En av de icke-lexikala kategorisymbolerna utgör dessutom grammatikens *startsymbol*, dvs. den symbol som får etikettera roten i ett analysträd.[11] Vi kommer i fortsättningen att anta att startsymbolen alltid är S.

Förrådet av *frasstrukturregler* är (enligt den här accepterade tolkningen) en uppsättning regler för konstruktion av etiketterade analysträd. Varje frasstrukturregel omfattar ett *vänsterled,* bestående av en icke-lexikal kategorisymbol, och ett *högerled,* bestående av en sekvens av kategorisymboler.[12]

Givet en frasstrukturgrammatik G (dvs. ett lexikon och en uppsättning frasstrukturregler enligt ovan) kan vi säga att G *tillåter* ett analysträd T om och endast om reglerna i G (inklusive den lexikala insättningsregeln) tillåter oss att konstruera T från en rot etiketterad med startsymbolen i G, dvs. givet en nod med symbolen S skall vi kunna konstruera resten av trädet

[10] Två mängder är *disjunkta* om och endast om de inte har några gemensamma element; se vidare t. ex. Allwood, Andersson & Dahl (1977) *Logic in Linguistics,* Cambridge University Press.

[11] Man kan också tänka sig att man tillåter flera olika startsymboler (som dock måste vara icke-lexikala kategorisymboler), men det vanligaste är att man antar en enda.

[12] Observera att högerledet kan innehålla såväl lexikala som icke-lexikala kategorisymboler medan vänsterledets kategorisymbol alltid måste vara icke-lexikal.

© Studentlitteratur

Kapitel 5 29 januari 1996 21.36:05 sida 46 av 48

enbart genom att följa regler i G. Därefter kan vi definiera mängden av *välformade analysträd* i förhållande till grammatiken G på följande sätt:

(55) Ett analysträd T är välformat i förhållande till en grammatik G om och endast om G tillåter T och T är terminerat.

Eftersom varje välformat träd kan ses som en representation av en sats (inklusive en bestämd konstituentstruktur), kan vi säga att en frasstrukturgrammatik som definierar en mängd välformade analysträd på detta sätt är en explicit redogörelse för en mängd av satser, dvs. en (generativ) grammatisk beskrivning av ett språk (jfr kapitel 1). Huruvida det också är en observationsmässigt och deskriptivt adekvat grammatisk beskrivning av det aktuella språket är naturligtvis en helt annan fråga.

© Studentlitteratur

Kapitel 5 29 januari 1996 21.36:05 sida 47 av 48

© Studentlitteratur

Kapitel 5 29 januari 1996 21.36:05 sida 48 av 48

6 Ett fragment av svenska

I stället för att göra anspråk på att ge grammatiska beskrivningar av hela språk, inskränker man sig ofta inom den generativa grammatiken till att formulera en grammatisk beskrivning av ett *fragment* av ett visst språk. Tanken bakom denna metodologi är att det är bättre att börja med att försöka ge en maximalt explicit och uttömmande redogörelse för en liten del av studieobjektet – en redogörelse som sedan successivt kan utsträckas till att omfatta allt mer – än att börja med en vag och ofullständig karakteristik av helheten (vilket är vad den s. k. traditionella grammatiken ofta anklagats för att göra). Utan att här ta ställning till metodens värde tänkte vi i alla fall ansluta oss till den generativa traditionen och formulera en frasstrukturgrammatik för ett litet fragment av svenska, dvs. för en liten delmängd av den totala mängden av svenska satser.

6.1 Syntaktiska kategorisymboler

De flesta av de syntaktiska kategorisymboler vi tänkte använda har förekommit tidigare i boken, men det kan kanske ändå vara bra att räkna upp dem och specificera deras avsedda innebörd (dvs. vilka kategorier de är avsedda att beteckna). Vi delar som vanligt upp kategorierna i lexikala och icke-lexikala:

Lexikala:	DET	Determinator (dvs. artiklar och ord med likartad syntaktisk funktion)
	N	Nomen (dvs. substantiv)
	A	Adjektiv
	PRO	Pronomen

© Studentlitteratur

	IV	Intransitivt verb
	TV	Transitivt verb
	P	Preposition
Icke-lexikala:	S	Sats
	NP	Nominalfras (dvs. en fras med ett nomen – eller pronomen – som huvudord)
	AP	Adjektivfras (dvs. en fras med ett adjektiv som huvudord)
	VP	Verbfras (dvs. en fras med ett verb som huvudord)
	PP	Prepositionsfras (dvs. en fras med en preposition som huvudord)

Dessutom bör vi specificera att S är grammatikens startsymbol, även om detta är ett standardantagande inom den generativa grammatiken.

6.2 Lexikon

Det lexikon vi tänkte använda är identiskt med det lexikon som återfinns i (52). För fullständighetens skull upprepar vi det här som (56).

© Studentlitteratur

Kapitel 6 29 januari 1996 21.36:20 sida 50 av 56

(56) DET: min
DET: en
DET: den
N: bror
N: cykeln
N: vän
N: damen
N: killen
N: handväskan
PRO: han
A: gammal
A: gamla
IV: sov
TV: köpte
TV: träffade
P: av
P: med

Dessutom måste vi ha en lexikal insättningsregel. Även här behåller vi den formulering vi tidigare gav i (53) och nu upprepar i (57).

(57) Om *a* är en lexikonenhet och *x* är en nod utan döttrar i ett analysträd, och om etiketten på *x* betecknar den lexikala kategori som *a* tillhör, så får du lov att lägga till en dotter till *x* med etiketten *a*.

6.3 Frasstrukturregler

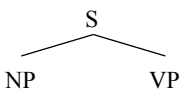
Till sist antar vi följande frasstrukturregler:

© Studentlitteratur

(58)
S -> NP VP
NP -> N
NP -> DET N
NP -> DET AP N
NP -> PRO
NP -> N PP
VP -> IV
VP -> TV NP
VP -> TV NP PP
AP -> A
PP -> P NP

Här finns en del nyheter, så ett par kommentarer kan nog vara på sin plats. Den första regeln har vi sett tidigare. I ”trädkonstruktiva” termer har den innebörden att en S-nod får ha en NP-nod följd av en VP-nod som döttrar. I mera allmängrammatiska termer säger den att en sats kan bestå av en nominalfras följd av en verbfras. Eftersom detta dessutom är den enda regeln med S som vänsterled, blir följden att alla satser som genereras av denna grammatik består av en nominalfras följd av en verbfras. Med andra ord kommer översta delen på alla välformade analysträd (i vårt lilla fragment) att se ut som i (59).

(59)



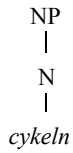
Om vi går vidare till nominalfrasreglerna (dvs. de regler som har symbolen NP som vänsterled), finner vi en hel rad av regler. Detta innebär alltså att nominalfraskonstituenter i vårt fragment (liksom i svenska i övrigt) kan ha olika struktur och att NP-noder i analysträd kan ha olika ”dotterskaror”. I (60–64) illustreras de olika NP-strukturer som tillåts av grammatiken till-

© Studentlitteratur

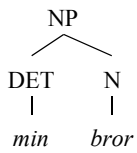
Kapitel 6 29 januari 1996 21.36:20 sida 52 av 56

sammans med exempel på nominalfraser som (enligt den provisoriska analys vi antar här) har respektive struktur.

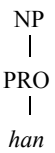
(60)



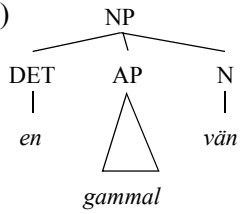
(61)



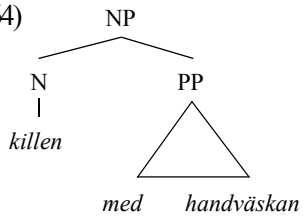
(62)



(63)

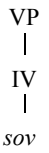


(64)

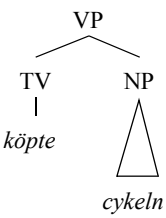


De tre verbfrasreglerna garanterar att VP-strukturerna i (65–67) kan förekomma i välformade analysträd.

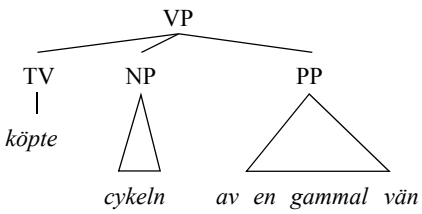
(65)



(66)



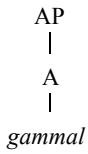
(67)



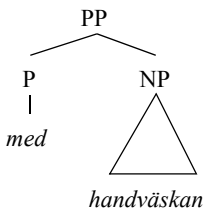
De två återstående reglerna ger oss AP-strukturen i (68) respektive PP-strukturen i (69).

© Studentlitteratur

(68)



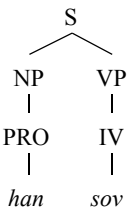
(69)



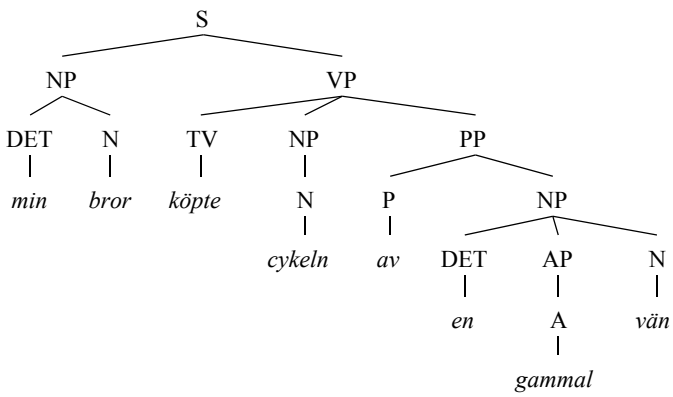
6.4 Exempel

Låt oss avsluta med att ge några exempel på analysträd som är välformade i förhållande till den grammatik vi just formulerat och som alltså representerar välformade svenska satser.

(70)



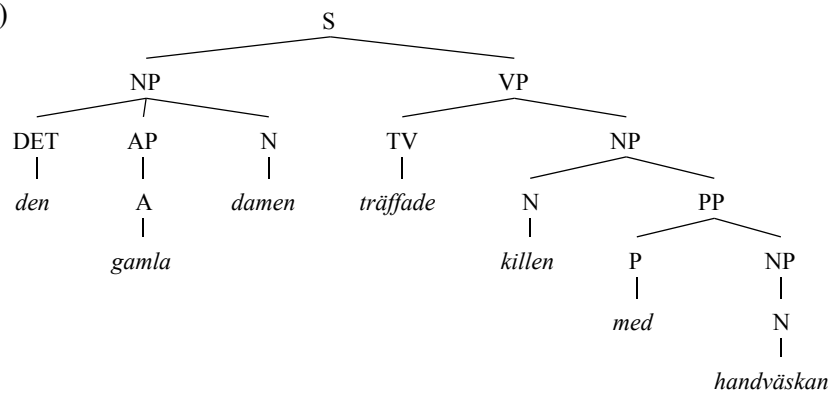
(71)



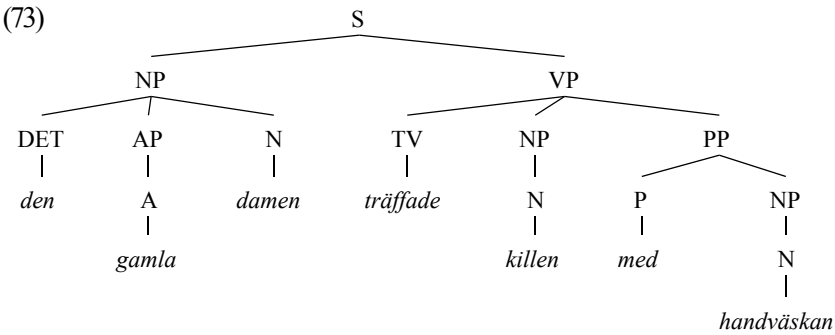
© Studentlitteratur

Kapitel 6 29 januari 1996 21.36:20 sida 54 av 56

(72)



(73)



Observera att (72) och (73) representerar två olika satser med samma (strukturellt tvetydiga) ordsträng (jfr avdelning 2.2).

6.5 Övningar

1. Betrakta analysträden (70–73). Avgör vilka av frasstrukturreglerna i (58) som är nödvändiga för konstruktionen av respektive träd.

2. Vilka av satserna (a-f) ingår i fragmentet av svenska som beskrivs i detta kapitel, dvs. vilka av ordsträngarna (a-f) kan förknippas med välformade analysträd enligt de regler som ges i (58).

© Studentlitteratur

Kapitel 6 29 januari 1996 21.36:20 sida 55 av 56

a. *min vän sov*
b. *handväskan träffade min bror*
c. *den gamla cykeln sov*
d. *killen köpte cykeln*
e. *killen köpte cykeln av den gamla damen*
f. *killen köpte cykeln av den gamla damen med handväskan*

3. Konstruera analysträd för de satser i övning 2 som ingår i det svenska fragmentet. Är någon av ordsträngarna strukturellt flertydig, dvs. kan någon av ordsträngarna förknippas med mer än ett välformat analysträd enligt de regler som ges i (58)? Stämmer detta resultat med dina intuitioner om satserna i fråga?

© Studentlitteratur

Kapitel 6 29 januari 1996 21.36:20 sida 56 av 56

7 Mer om frasstrukturgrammatik

I kapitel 5 och 6 har vi introducerat en särskild modell för grammatiska beskrivningar – frasstrukturgrammatik – och visat hur den kan användas för att beskriva ett litet fragment av svenska. I detta kapitel skall vi först introducera några förkortningskonventioner som underlättar arbetet med att skriva frasstrukturgrammatiker. Dessa konventioner används flitigt i litteraturen och är därför bra att känna till. I den senare delen av kapitlet skall vi sedan utvidga vår beskrivning av svenska och diskutera några mer komplexa grammatiska strukturer och deras behandling inom frasstrukturgrammatik.

7.1 Förkortningskonventioner

När man skriver frasstrukturregler använder man ofta ett antal förkortningskonventioner som bidrar till att minska det totala antalet regler i grammatiken. De två viktigaste berör *optionella* respektive *alternativa* konstituenter.

Låt oss återvända till nominalfrasreglerna i det svenska fragmentet, som återfinns i (74).[1]

(74) NP -> N
NP -> DET N
NP -> DET AP N
NP -> N PP

Den första regeln säger att en nominalfras kan bestå av ett ensamt nomen, och den andra att en nominalfras kan bestå av en determinator följd av ett nomen. Ett sätt att sammanfatta innebörden i *båda* reglerna skulle därför

[1] Vi bortser i detta sammanhang från regeln NP -> PRO.

© Studentlitteratur

vara att säga att en nominalfras alltid består av ett nomen och att detta nomen kan (men inte måste) föregås av en determinator, eller – för att använde den gängse terminologin – att en nominalfras har ett nomen som *obligatorisk* konstituent och en föregående determinator som *optionell* (eng. *optional)* eller *frivillig* konstituent. För att kunna uttrycka generaliseringar av denna typ har man infört konventionen att optionella konstituenter får skrivas inom parentes i frasstrukturregler. Med hjälp av denna konvention kan vi alltså ersätta de båda reglerna ovan med den enda regeln (75).

(75) NP -> (DET) N

Låt oss nu betrakta även den tredje nominalfrasregeln i (74), som säger att en nominalfras kan bestå av en determinator, följd av en adjektivfras, följd av ett nomen. Om vi jämför denna regel med de båda föregående, så ser vi att nomenet alltjämt är en obligatorisk konstituent i nominalfrasen medan vi nu har två optionella konstituenter, nämligen determinatorn och adjektivfrasen. De två optionella konstituenterna har dock inte riktigt samma status, eftersom determinatorn kan förekomma såväl med (jfr tredje regeln) som utan adjektivfrasen (jfr andra regeln), medan adjektivfrasen enbart kan förekomma om determinatorn också gör det. Detta innebär bl. a. att vi *inte* kan sammanfatta de tre första nominalfrasreglerna på följande sätt:

(76) NP -> (DET) (AP) N

En regel som (76) säger nämligen att vi kan ha vilka kombinationer som helst av de tre kategorierna i högerledet så länge bara N är närvarande, dvs. (76) är en förkortning av följande uppsättning regler:

(77) NP -> N
NP -> DET N
NP -> AP N
NP -> DET AP N

Den tredje regeln representerar här en nominalfrasstruktur som vi inte hade i det ursprungliga fragmentet.

Den fjärde och sista av de ursprungliga nominalfrasreglerna säger att en nominalfras kan bestå av ett nomen följt av en prepositionsfras. Eftersom även prepositionsfrasen måste betraktas som en optionell konstituent (den återfinns ju inte i någon av de övriga reglerna), skulle vi kunna formulera följande generella regel för nominalfrasers struktur:

(78) NP -> (DET) (AP) N (PP)

© Studentlitteratur

Kapitel 7 21 december 2009 18.25:20 sida 58 av 66

Denna regel avviker dock ännu mer från den ursprungliga regeluppsättningen än vad (76) gör, eftersom (78) är en förkortning för följande regler:

(79) NP -> N
NP -> DET N
NP -> AP N
NP -> N PP
NP -> DET AP N
NP -> DET N PP
NP -> AP N PP
NP -> DET AP N PP

Regeluppsättningen i (79) innehåller fyra regler (den tredje, sjätte, sjunde och åttonde) som inte finns i den ursprungliga. Låt oss därför slå fast att det enda sättet att förkorta (74) är i form av följande regler:[2]

(80) NP -> N (PP)
NP -> DET (AP) N

På liknande sätt kan två av de ursprungliga verbfrasreglerna i (81) förkortas som i (82).

(81) VP -> TV NP
VP -> TV NP PP

(82) VP -> TV NP (PP)

Vi nämnde inledningsvis att det också finns en annan typ av förkortningskonvention, som används i samband med alternativa konstituenter. För att kunna illustrera denna konvention, måste vi emellertid först införa några nya regler.

På svenska finns det många verb som kan bilda en verbfras antingen tillsammans med en nominalfras (ett direkt objekt) eller med en prepositionsfras (ett prepositionsobjekt). Vi har t. ex. *läsa en bok* och *läsa i en bok,* liksom *skriva en bok* och *skriva i en bok* (eller *på en bok)*. För att kunna beskriva dessa verbfrasstrukturer skulle vi alltså behöva följande regler (varav endast den första återfinns i vårt fragment):

[2] Observera att vad som diskuteras här inte är huruvida (76) och (78) är bättre beskrivningar av svenskans nominalfraser än (74), utan enbart det tekniska problemet att förkorta den ursprungliga regeluppsättningen. Det är dessutom helt klart att ingen av beskrivningarna i detta kapitel är en helt adekvat beskrivning av strukturen hos svenska nominalfraser.

© Studentlitteratur

Kapitel 7 21 december 2009 18.25:20 sida 59 av 66

(83) VP -> TV NP
VP -> TV PP

Om vi försöker sammanfatta innebörden i båda reglerna, kan vi säga att en verbfras kan bestå av ett transitivt verb (obligatorisk konstituent), följt av antingen en nominalfras eller en prepositionsfras (men inte båda eller ingendera). Nominalfrasen och prepositionsfrasen är med andra ord alternativa konstituenter i verbfrasen. De är däremot inte optionella, eftersom en av dem måste vara med. Konventionen för att markera alternativa konstituenter är att innesluta dem i mängdklamrar (jfr exempel 10 i kapitel 1). Reglerna i (83) kan således förkortas på följande sätt:

(84) VP -> TV {NP, PP}

Observera att ordningsföljden mellan de alternativa konstituenterna i regeln inte har någon betydelse, eftersom de aldrig kan förekomma tillsammans.[3] Detta innebär att (85) är helt ekvivalent med (84).

(85) VP -> TV {PP, NP}

Ett alternativ kan också bestå av en sekvens av kategorier. T. ex. kan reglerna i (86) förkortas som i (87):

(86) VP -> TV NP
VP -> TV PP
VP -> TV NP PP

(87) VP -> TV {NP, PP, NP PP}

Observera att man i detta fall måste vara noga med att *inte* sätta kommatecken inuti den sekvens av kategorier som tillsammans utgör ett alternativ (dvs. NP PP).[4]

Det bör också nämnas att de två förkortningskonventionerna kan förekomma i en och samma regel såsom i (89), vilket utgör en förkortning av (88).

[3] Detta är också anledningen till att man valt just mängdklamrar som notation, eftersom elementen i en mängd inte heller är ordnade inbördes.

[4] Denna notationsvariant är på sätt och vis mer överskådlig (och man slipper risken med att sätta kommatecken på fel ställe) men den har nackdelen att vara icke-linjär, vilket gör den svår att (re)producera i många sammanhang.

© Studentlitteratur

Kapitel 7 21 december 2009 18.25:20 sida 60 av 66

(88) VP -> TV NP
VP -> TV PP
VP -> TV NP PP
VP -> TV PP PP

(89) VP -> TV {NP, PP} (PP)

De två konventionerna kan också kombineras med varandra så som i (91), vilket utgör en förkortning av (90).

(90) VP -> IV
VP -> IV PP
VP -> IV ADVP

(91) VP -> IV ({PP, ADVP})

Även för lexikonet finns möjlighet till förkortningar. En vanlig modell är att man räknar upp alla ord som tillhör en viss lexikal kategori på en gång, t. ex. på följande sätt (jfr 56 ovan):

(92) DET: {min, en, den}
N: {bror, cykeln, vän, damen, killen, handväskan}
PRO: {han}
A: {gammal, gamla}
IV: {sov}
TV: {köpte, träffade}
P: {av, med}

7.2 Rekursiva strukturer

I kapitel 1 nämnde vi mer eller mindre i förbigående att mängden av satser i ett naturligt språk vanligen antas vara oändlig, eftersom det inte finns någon principiell begränsning på hur lång en sats kan vara. Av detta följer att en adekvat grammatisk beskrivning av ett naturligt språk måste definiera (eller generera) ett oändligt antal välformade analysträd.

En förutsättning för att en grammatik (med ett ändligt antal regler) skall kunna definiera en oändlig mängd analysträd är emellertid att den tillåter *rekursiva strukturer.* Att en struktur är rekursiv innebär att den innehåller en delstruktur av samma typ som helheten. Mera specifikt i samband med ana-

© Studentlitteratur

lysträd kan vi säga att en trädstruktur är rekursiv om den innehåller en nod som dominerar en annan nod med samma etikett.

Det enklaste sättet att åstadkomma rekursiva strukturer i en frasstrukturgrammatik är att införa *intrinsikalt rekursiva regler,* dvs. regler vars vänsterled också återfinns i högerledet. Ett exempel på en sådan regel är (93).

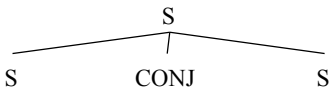
(93) S -> S CONJ S

Denna regel säger att en S-nod kan ha tre döttrar, varav den första och den sista har etiketten S, medan ”mellandottern” har etiketten CONJ. CONJ (eng. *conjunction)* är en ny lexikal kategorisymbol som betecknar en kategori som ungefär motsvarar den traditionella kategorin *samordnande konjunktioner.* Antag nu att vi utökar det svenska fragmentet genom att lägga CONJ till förrådet av kategorisymboler och (93) till förrådet av frasstrukturregler, samt utökar lexikonet med följande enheter:

(94) CONJ: och
CONJ: men

I kraft av regel (93) kan toppen på välformade trädstrukturer nu se ut som i (95), vilket alltså utgör ett alternativ till den gamla vanliga NP-VP-strukturen.

(95)



Under CONJ-noden kan vi sedan sätta in *och* eller *men*, men frågan är vad som gäller för de båda S-noderna? Svaret är att de kan dominera precis samma sak som vilken annan S-nod som helst. Det faktum att de två dotternoderna i (95) inte utgör rotnoder har med andra ord ingen som helst betydelse för hur man kan bygga vidare på dem. Varje regel som har symbolen S som vänsterled kan tillämpas på varje nod som har symbolen S som etikett, oavsett var i ett träd noden befinner sig. Innebörden av regel (93) blir med andra ord att en sats kan bestå av en sats, följd av en samordnande konjunktion, följd av en sats.[5]

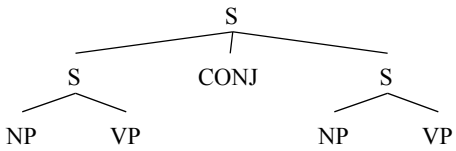
[5] En sådan sats utgör en *rekursiv struktur* enligt den allmänna karakteristiken ovan, dvs. den innehåller en delstruktur (i det här fallet t. o. m. två) av samma typ som helheten. Ett analysträd som representerar en sådan sats utgör en *rekursiv trädstruktur,* eftersom den innehåller en nod (rotnoden) som dominerar en nod (i det här fallet t. o. m. två) med samma etikett.

© Studentlitteratur

Kapitel 7 21 december 2009 18.25:20 sida 62 av 66

Detta innebär att vi t.ex. kan tillämpa den traditionella NP-VP-regeln på båda S-döttrarna och få en struktur av följande utseende:

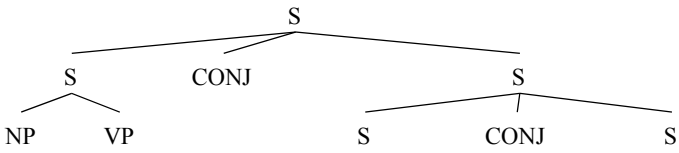
(96)



Denna struktur kan sedan utvecklas vidare till analysträd motsvarande satser som *den gamla damen sov och killen köpte cykeln* och *min bror träffade killen men han sov.*

Vi kan emellertid lika gärna tillämpa den nya regeln (93) på en av S-döttrarna (eller båda). Detta kan ge en struktur som (97), som sedan kan byggas ut till analysträd för satser som t. ex. *den gamla damen köpte cykeln och min bror träffade killen men han sov.*

(97)



En frasstrukturgrammatik som tillåter rekursiva strukturer definierar ett oändligt antal välformade analysträd, eftersom det inte finns någon principiell begränsning på hur många gånger vi kan införa en nod som domineras av en nod med samma etikett och därmed så att säga ”börja om från början” men på en lägre nivå. Om vi utökar det svenska fragmentet på det sätt vi föreslagit i denna avdelning, får vi följaktligen en grammatik som genererar ett oändligt antal satser, även om den fortfarande bara genererar ett fragment av alla satser i svenska.

En grammatik måste emellertid inte innehålla intrinsikalt rekursiva regler (dvs. regler vars vänsterled återfinns i högerledet) för att tillåta rekursiva strukturer. En annan möjlighet är att flera regler i samverkan möjliggör rekursiva strukturer.

Antag till exempel att vi utökar det svenska fragmentet genom att införa tre nya kategorisymboler: S', COMP och PV. Den första (som utläses ”S prim” eller ”S bar”[6]) är en icke-lexikal kategorisymbol som är avsedd att beteckna

© Studentlitteratur

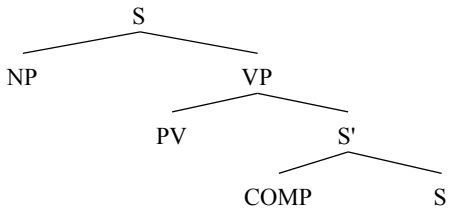
den syntaktiska kategorin *bisatser.* De två senare är lexikala kategorisymboler som är avsedda att beteckna dels en kategori som ungefär motsvaras av den traditionella kategorin *underordnande konjunktioner* (COMP efter engelskans *complementizer),* dels den undergrupp av verb som tar *att*-satser som objekt (PV för *propositionellt verb*). Antag vidare att vi utökar lexikonet med enheterna i (98) och förrådet av frasstrukturregler med reglerna i (99).

(98) COMP: att
PV: sa

(99) VP -> PV S'
S' -> COMP S

Ingen av reglerna i (99) är intrinsikalt rekursiv, men tillsammans med de ursprungliga frasstrukturreglerna i fragmentet tillåter de rekursiva strukturer av följande utseende:

(100)



Trädet i (100) är en rekursiv trädstruktur, eftersom rotnoden (med etiketten S) dominerar en annan nod med samma etikett. En struktur av denna typ passar för satser som t. ex. *killen sa att min bror köpte cykeln,* där *min bror köpte cykeln* är en (struktur av kategorin) S, *att min bror köpte cykeln* en S', och *sa att min bror köpte cykeln* en VP.

I introduktionsböcker till generativ grammatik finner man ibland påståendet att en grammatik måste innehålla rekursiva regler för att generera ett oändligt antal satser. Om man med ”rekursiva regler” menar *intrinsikalt rekursiva regler,* är detta påstående som vi har sett falskt. Om man däremot tolkar ”rekursiva regler” som *regler som (ensamma eller tillsammans) tillåter rekursiva strukturer,* är det sant (men formulerat på ett vilseledande sätt).

[6] Beteckningen ”S bar” härrör från den alternativa notationen $\overline{S}$, som på engelska utläses som ”S bar”.

© Studentlitteratur

Kapitel 7 21 december 2009 18.25:20 sida 64 av 66

7.3 Övningar

1. Försök att förkorta följande uppsättning frasstrukturregler så mycket som möjligt genom att använda konventionerna för optionella och alternativa konstituenter.

   S -> NP VP
   S -> NP VP PP
   NP -> N
   NP -> DET N
   NP -> AP N
   NP -> N PP
   NP -> DET N PP
   NP -> AP N PP
   VP -> TV PP
   VP -> TV NP
   VP -> TV NP PP

2. Försök att formulera om följande uppsättning frasstrukturregler utan att använda några förkortningskonventioner.

   S -> NP {VP, AP}
   NP -> DET (AP) N (PP)
   VP -> DV NP {NP, PP} (PP)
   AP -> A (PP)
   PP -> P {NP, AP}

3. Försök att utöka det svenska fragmentet från kapitel 6 så att även följande satser (som du jobbat med i flera tidigare övningar) genereras.

   a. *Pelle älskar Lisa*
   b. *Pelle älskar sin fru*
   c. *katten stirrade på råttan*
   d. *katten stirrade på råttan och råttan stirrade på katten*
   e. *katten och råttan stirrade på varandra*
   f. *jag gav honom en bok*
   g. *jag gav min tjatiga kusin en bok om grammatik*
   h. *det nya kaféet ligger på hörnet*
   i. *Pelle kysste flickan i kiosken*

© Studentlitteratur

Du måste förmodligen göra tillägg såväl i lexikonet som i uppsättningarna av kategorisymboler och frasstrukturregler. Eventuellt kan du ha hjälp av de utvidgningar som diskuteras i detta kapitel. Tänk också på att (den strukturellt flertydiga) ordsträngen i (i) bör förknippas med (åtminstone) två analysträd.

4. Den grammatik du formulerat i övning 3 (som en utvidgning av vårt svenska fragment) genererar en mängd av satser (representerade av välformade analysträd). Är denna mängd ändlig eller oändlig?

© Studentlitteratur

Kapitel 7 21 december 2009 18.25:20 sida 66 av 66

8 Bortom frasstrukturgrammatik

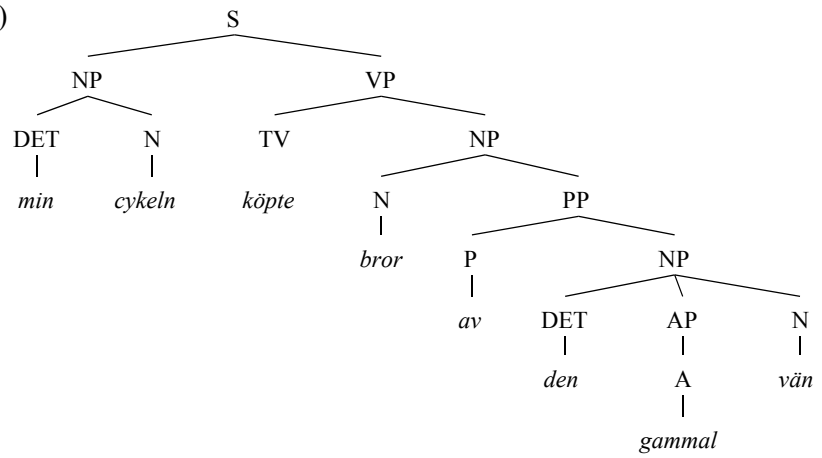
Denna bok har till övervägande delen handlat om frasstrukturgrammatik som en modell för grammatisk beskrivning av naturliga språk. I detta avslutande kapitel skall vi säga något kort om de begränsningar som denna beskrivningsmodell har och om de åtgärder som föreslagits för att överkomma dessa begränsningar.

8.1 Problem med frasstrukturgrammatiska beskrivningar

I kapitel 1 konstaterade vi att en grammatisk beskrivning av ett språk (t. ex. svenska) är observationsmässigt adekvat om och endast om den för det första genererar alla satser i språket och för det andra genererar endast satser i språket. Den frasstrukturgrammatik vi formulerade i kapitel 6 och utvidgade i kapitel 7 genererar utan tvekan svenska satser, dvs. den definierar välformade analysträd som representerar svenska satser, men den genererar absolut inte *alla* svenska satser (utan endast ett *fragment* av mängden av svenska satser). Detta är emellertid inte så allvarligt, eftersom tanken är att en beskrivning av ett språkfragment kan utvidgas till att omfatta allt större delar av språket och till sist (i bästa fall) hela språket. Vad som är betydligt värre är emellertid att vår grammatik inte genererar *endast* svenska satser, dvs. att den genererar analysträd som inte representerar svenska satser alls. Betrakta t. ex. följande analysträd, som är välformat enligt den frasstrukturgrammatik vi formulerade i kapitel 6.

© Studentlitteratur

(101)



Låt oss betrakta de presumtiva nominalfraserna *min cykeln, bror* och *den gammal vän*. Det är alldeles klart att dessa ordkombinationer *inte* är välformade svenska nominalfraser, och problemet tycks vara att de frasstrukturregler vi formulerat för svenska nominalfraser (dvs. de regler som har NP som vänsterled) inte är tillräckligt restriktiva utan tillåter för mycket. De *övergenererar* för att använda den gängse termen. Det är helt enkelt inte så att vilket nomen som helst ensamt kan utgöra en nominalfras på svenska *(cykeln* och *han* kan det, men inte *bror* och *vän*), eller att vilken determinator som helst kan kombineras med vilket nomen som helst till en välformad nominalfras *(min bror* och *en vän* går bra, men *min cykeln* och *en killen* går inte alls). På samma sätt kan inte vilken adjektivfras som helst stoppas in mellan en determinator och ett nomen *(den gamla damen* och *en gammal vän* är välformade, men *den gammal damen* och *en gamla vän* är det inte).

Vi skall inte här fördjupa oss i de olika förslag som har framförts för att komma till rätta med övergenereringsproblemet, men det är ett problem som starkt bidragit till att forma teoriutvecklingen inom det generativa paradigmet. Det grundläggande problemet kan sägas vara att vår kategorisering av substantiv, adjektiv och determinatorer hittills varit för grovkornig. För att kunna filtrera bort icke välformade nominalfraser som *bror, min cykeln*, *den gammal vän* osv. måste vi skilja på *bestämda* och *obestämda* substantiv (och adjektiv och determinatorer). I ett större perspektiv måste vi också skilja på substantiv i *singular* och *plural*, substantiv som har genus *utrum* och substantiv som har genus *neutrum*, substantiv som står i *grundform* och substantiv som står i *genitiv* osv. Vad värre är: Alla dessa underka-

© Studentlitteratur

Kapitel 8 29 januari 1996 21.15:36 sida 68 av 72

tegorier kan kombineras med varandra på olika sätt, vilket gör att vi måste ha en kategori för *neutrala substantiv i grundform singular*, en annan för *neutrala substantiv i grundform plural* osv. Man inser lätt att det kommer att krävas ett mycket stort antal kategorier, och därmed i sin tur ett mycket stort antal regler, för att ge en någorlunda heltäckande beskrivning av t. ex. svenska.

Inom de flesta moderna grammatiska teorier har man därför övergivit den rena frasstrukturgrammatikens mycket enkla kategorisymboler, där man i princip antar en distinkt symbol för varje kategori som behöver urskiljas. I stället använder man sig av s. k. *komplexa kategorier* uppbyggda av *syntaktiska särdrag*. Med denna begreppsapparat blir det möjligt att säga att t. ex. *cykel* och *cykeln* tillhör samma kategori (dvs. nomen) på en viss abstraktionsnivå men skiljer sig åt genom att ha olika *särdrag* (egenskaper) med avseende på bestämdhet. Information om dessa ords kategorier kan t. ex. representeras på följande sätt i två s. k. *särdragsstrukturer*:

(102) *cykel*

$$\begin{bmatrix} \text{CAT} & \text{N} \\ \text{GEN} & \text{UTER} \\ \text{CASE} & \text{NOM} \\ \text{NUM} & \text{SING} \\ \text{DEF} & - \end{bmatrix}$$

cykeln

$$\begin{bmatrix} \text{CAT} & \text{N} \\ \text{GEN} & \text{UTER} \\ \text{CASE} & \text{NOM} \\ \text{NUM} & \text{SING} \\ \text{DEF} & + \end{bmatrix}$$

Ett annat problem med frasstrukturgrammatiska beskrivningar som haft stor betydelse för den generativa grammatikens utveckling är problemet med diskontinuerliga konstituenter (jfr avdelning 2.2). Om övergenereringsproblemet handlar om att grammatiken tillåter för mycket (och därför inte genererar *endast* grammatikaliska satser), så har problemet med diskontinuerliga konstituenter i stället att göra med att grammatiken tillåter för litet (och därför inte genererar *alla* grammatikaliska satser). Betrakta följande ordsträngar:

(103) *min bror köpte cykeln*
köpte min bror cykeln

© Studentlitteratur

Kapitel 8 29 januari 1996 21.15:36 sida 69 av 72

Det råder ingen tvekan om att båda ordsträngarna i (103) representerar grammatikaliska satser i svenska. Dessutom är det uppenbart att det finns ett nära samband mellan dem. Speciellt ligger det nära till hands (som vi konstaterade i avdelning 2.2) att tillskriva dem samma konstituentstruktur. Problemet är emellertid att detta skulle innebära (om vi håller fast vid den traditionella analysen av den första satsen) att *köpte* och *cykeln* skulle utgöra en konstituent (närmare bestämt en verbfras) i den andra satsen. Men denna konstituent skulle då vara *diskontinuerlig* (dvs. den skulle inte motsvaras av någon sammanhängande delsträng av den totala ordsträngen), och som den klassiska frasstrukturgrammatiken ser ut finns det ingen möjlighet att generera analysträd med sådana konstituenter.[1]

Problemet med diskontinuerliga konstituenter är i likhet med övergenereringsproblemet ett problem som haft ett avgörande inflytande på teoriutvecklingen inom det generativa paradigmet. Särskilt har det motiverat införandet av s. k. *transformationsregler* (dvs. regler som bl. a. kan flytta konstituenter i analysträd). Denna typ av regler har varit viktiga framför allt inom de teorier som föreslagits av Noam Chomsky.

8.2 Frasstrukturgrammatikens roll i modern grammatisk teori

Som vi antytt flera gånger spelar frasstrukturgrammatiken i sin klassiska form ingen framträdande roll i dagens generativa grammatiska teorier. Det finns emellertid flera goda skäl att ändå ägna frasstrukturgrammatiken mer än bara en flyktig kommentar. Låt oss avslutningsvis nämna tre av dessa.

För det första är de *syntaktiska analysträd* som spelar en central roll inom så gott som alla moderna teorier väsentligen identiska med de syntaktiska analysträd som man finner inom frasstrukturgrammatiken förutom att man numera oftast ersätter de enkla kategorisymbolerna med komplexa särdragsstrukturer (det är i första hand reglerna som definierar träden som förändrats genom åren).

[1] Om analysträd skall kunna representera diskontinuerliga konstituenter, måste de kunna innehålla ”korsande grenar”, och detta är omöjligt i den klassiska frasstrukturgrammatiken.

© Studentlitteratur

Kapitel 8 29 januari 1996 21.15:36 sida 70 av 72

För det andra finns *frasstrukturregler* i någon form kvar i många moderna teorier, även om de bara utgör en liten del av den totala beskrivningsapparaten. Man har med andra ord ofta *kompletterat* eller *modifierat* frasstrukturreglerna (och gett dem en mindre framträdande roll) snarare än *ersatt* dem.

För det tredje har frasstrukturgrammatiken, som vi antydde i föregående avsnitt, spelat en viktig roll i den generativa grammatikens utveckling i den meningen att många av de nya begrepp och metoder som införts har kommit till för att lösa problem som ursprungligen uppkom i samband med frasstrukturgrammatiska beskrivningar (t. ex. övergenereringsproblemet och problemet med diskontinuerliga konstituenter). En ordentlig förståelse av den klassiska frasstrukturgrammatiken och dess problem är därför en viktig förutsättning för en djupare förståelse av den utveckling som lett fram till dagens generativa grammatiska teorier. Vi hoppas att denna bok skall kunna lägga grunden till denna förståelse.

© Studentlitteratur

Kapitel 8 29 januari 1996 21.15:36 sida 71 av 72

© Studentlitteratur

Kapitel 8 29 januari 1996 21.15:36 sida 72 av 72

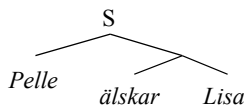
Lösningsförslag till övningar

I nedanstående lösningsförslag finns ibland mer än en analys av en och samma sats. Detta kan bero på att det finns mer än en uppfattning hos experterna om hur en viss konstruktion skall analyseras eller att konstruktionen ännu inte analyserats i detalj. Vi har dock inte gjort något försök att få med alla tänkbara analyser av varje sats, vilket gör att det kan finnas fullt rimliga analyser som inte återfinns bland lösningsförslagen. I de fall vi funnit det motiverat har vi dock kommenterat alternativa analyser och förklarat vari skillnaden består.

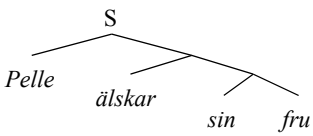
Kapitel 2: Satsers struktur

Övning 1

(a)

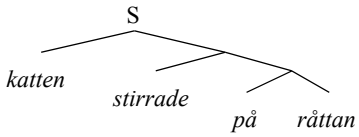


(b)

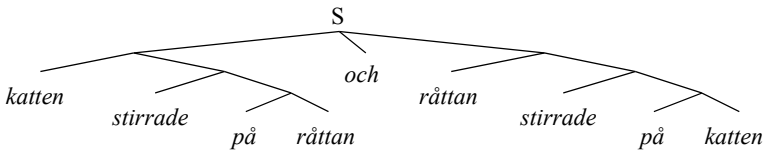


© Studentlitteratur

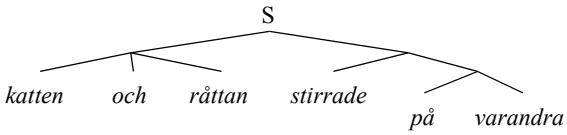
(c)



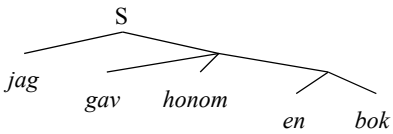
(d)



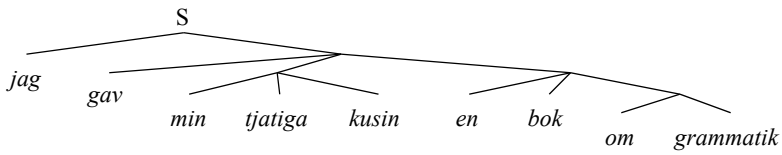
(e)

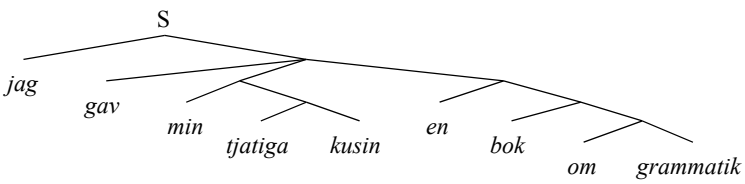


(f)



(g)

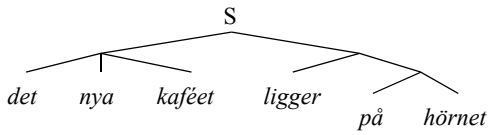




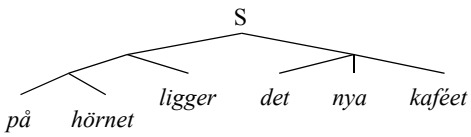
© Studentlitteratur

Kommentar: Den enklaste analysen av treledade fraser som *min tjatiga kusin* och *en bok om grammatik* är att låta dem bestå av tre omedelbara konstituenter: *min – tjatiga – kusin* respektive *en – bok – om grammatik*. Detta är också den analys som vi håller oss till i denna bok och som återges i det första analysträdet. De flesta teoretiker anser emellertid att en mera korrekt analys är att först dela fraserna i två delar (*min – tjatiga kusin* respektive *en – bok om grammatik*), såsom vi gjort i det andra trädet. Att gå in på argumenten för den senare analysen skulle emellertid föra alltför långt i en bok som denna. (Vi återkommer till analysen av dessa fraser i övning 1 i kapitel 3.)

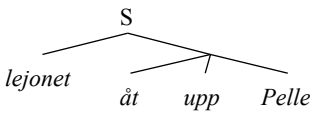
(h)

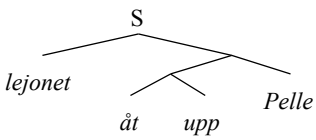


(i)



(j)



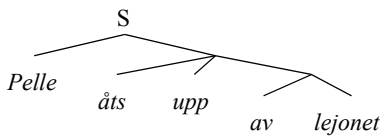


Kommentar: Ett uttryck som *åt upp* kallas ofta *partikelverb*, eftersom det består av ett verb (*åt*) och en partikel (*upp*). Om man betraktar detta som ett enda ord (trots att det skrivs med mellanrum), kan det förefalla naturligt att dela frasen *åt upp Pelle* i två delar (*åt upp – Pelle*), såsom i det andra trädet. Här kommer vi dock att hålla oss till den första analysen, där *åt*, *upp* och

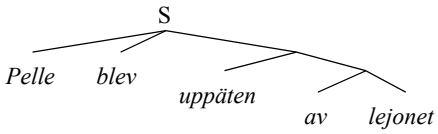
© Studentlitteratur

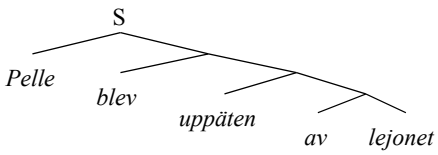
Pelle alla ses som omedelbara konstituenter av frasen *åt upp Pelle*. (Samma resonemang gäller för nästföljande sats.)

(k)



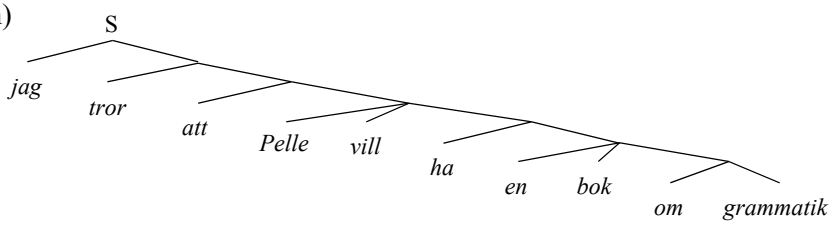
(l)





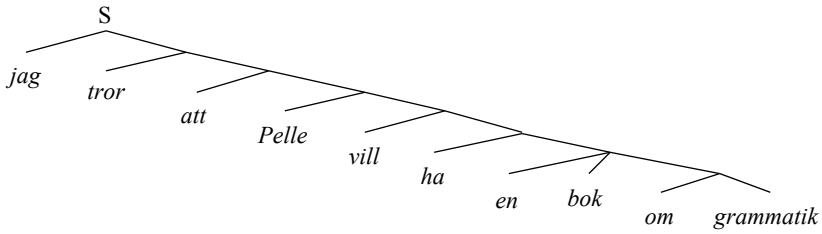
Kommentar: Skillnaden mellan de två analyserna beror i detta fall på vilken status man tilldelar hjälpverbet *blev* i förhållande till andra verb. Vi återkommer till detta i övning 1 i kapitel 3. (I det följande kommer vi dock att hålla oss till den första analysen, där hjälpverbet ses som en omedelbar konstituent av satsen. Samma sak gäller för hjälpverbet *vill* i nästföljande sats.)

(m)

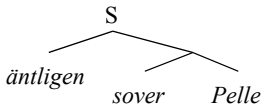


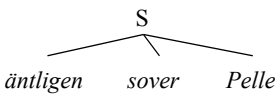
© Studentlitteratur

Bilaga A 29 January 1996 21.35:51 sida 84 av 106

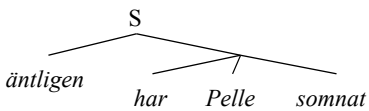


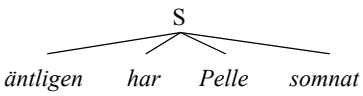
(n)





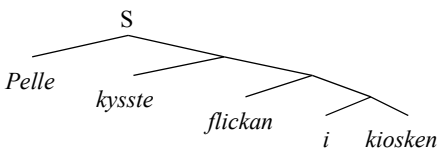
(o)





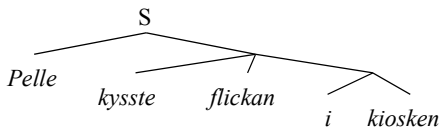
Övning 2

(a)



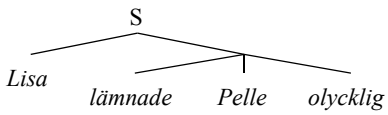
© Studentlitteratur

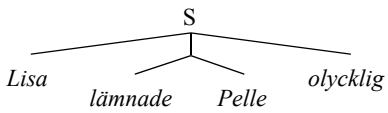
.



Kommentar: Tvetydigheten är i detta fall av samma slag som i exemplet *den gamla damen träffade killen med handväskan*, dvs. den beror på att frasen *i kiosken* kan tolkas antingen som attribut till substantivet *flickan* eller som (rums)adverbial.

(b)

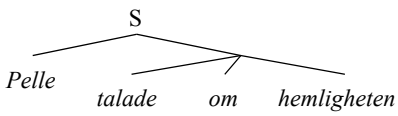




Kommentar: Här kan det s. k. predikativa attributet *olycklig* antingen vara en bestämning till *Pelle* (dvs. Pelle var olycklig då Lisa lämnade honom) eller till *Lisa* (dvs. Lisa var olycklig då hon lämnade Pelle). Det är inte alldeles klart vilken konstituentstruktur dessa respektive tolkningar har, men ovan ges i alla fall två förslag.

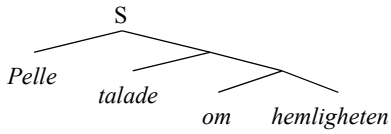
(c) Tvetydigheten har i detta fall att göra med huruvida det måste vara samma två språk som alla personer talar. Denna tvetydighet kan dock knappast förklaras som strukturell.

(d)



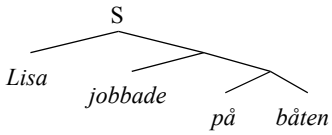
© Studentlitteratur

Bilaga A 29 January 1996 21.35:51 sida 86 av 106



Kommentar: Här är frågan huruvida *om* är en partikel som hör ihop med verbet *talade* (jfr 1j ovan) eller en preposition som hör ihop med objektet *hemligheten*.

(e)



Kommentar: Frasen *på båten* kan här tolkas antingen som s. k. prepositionsobjekt (jfr *Lisa jobbade på en artikel*) eller som rumsadverbial (jfr *Lisa jobbade på havet*). Denna tvetydighet är således av grammatisk natur, men det är inte klart att den kan förklaras enbart i termer av konstituentstruktur. Med de antaganden vi gjort hittills får de två tolkningarna samma struktur.

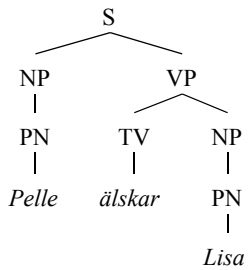
© Studentlitteratur

Bilaga A 29 January 1996 21.35:51 sida 87 av 106

Kapitel 3: Syntaktiska kategorier

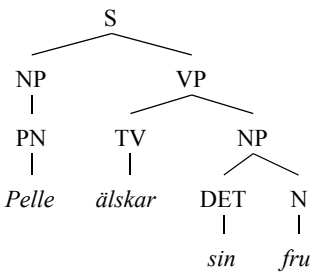
Övning 1

(a)



Kommentar: Uttryck som *Pelle* och *Lisa* är exempel på *egennamn*, vilka i traditionell grammatik brukar räknas som en sorts substantiv. Det finns emellertid goda skäl att tilldela dem en egen kategori, som ofta får symbolen PN (efter engelskans *proper name*).

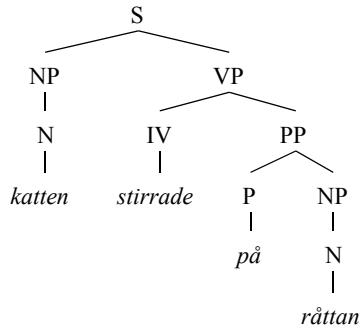
(b)



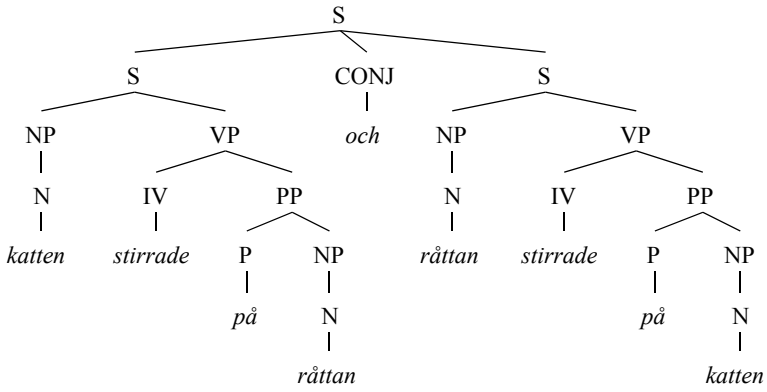
© Studentlitteratur

Bilaga A 29 January 1996 21.35:51 sida 88 av 106

(c)



(d)

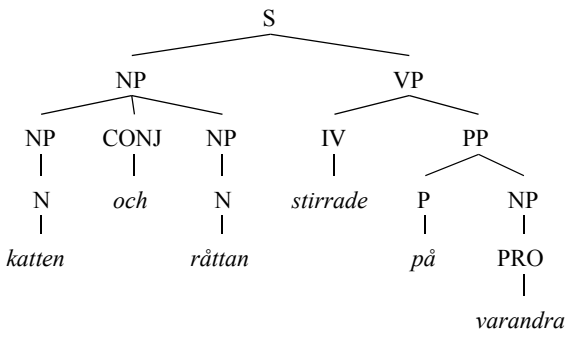


Kommentar: Här behöver vi en ny kategori för *samordnande konjunktioner* såsom *och*. Vi använder kategorisymbolen CONJ (efter engelskans *conjunction*). (Jfr också kapitel 7.)

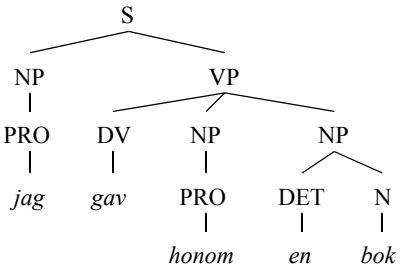
© Studentlitteratur

Bilaga A 29 January 1996 21.35:51 sida 89 av 106

(e)

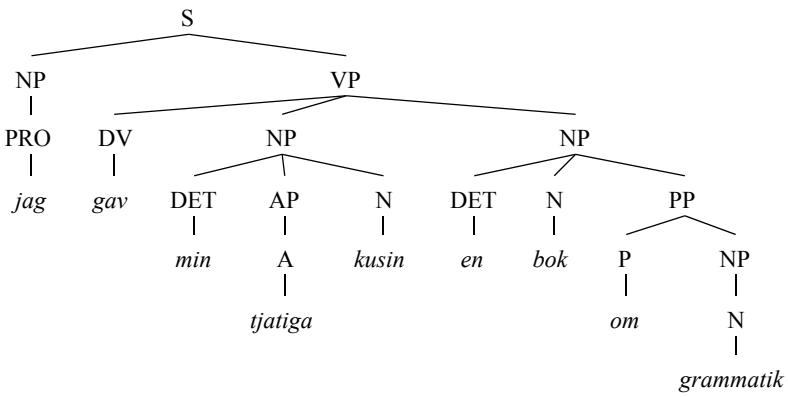


(f)



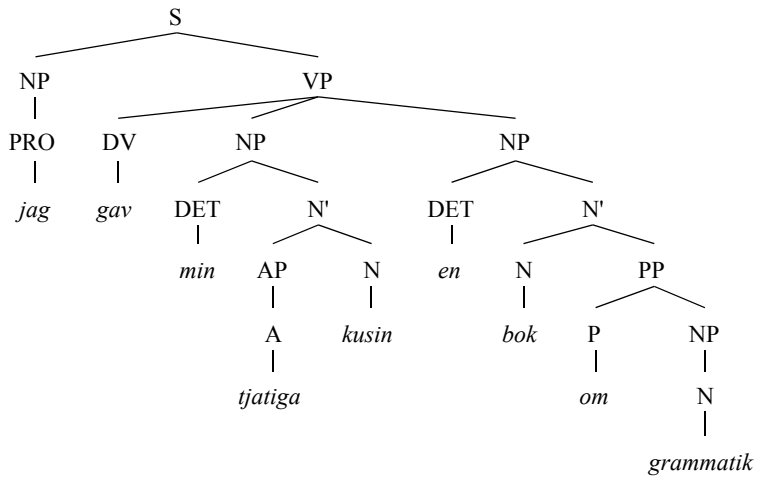
Kommentar: Verbet *gav* skiljer sig från de verb vi sett hittills genom att det tar två objekt. Sådana verb kallas *ditransitiva verb* (DV).

(g)



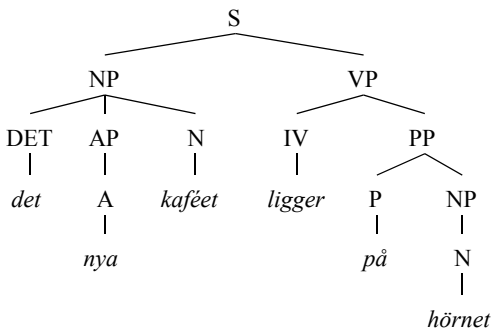
© Studentlitteratur

Bilaga A 29 January 1996 21.35:51 sida 90 av 106



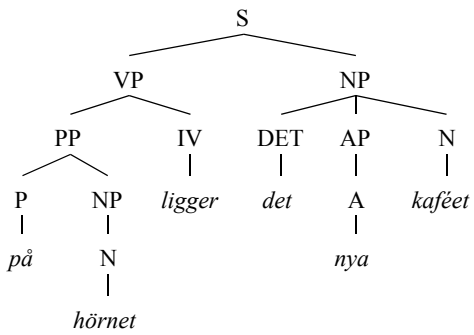
Kommentar: Om vi vill anta den mer hierarkiska analysen av nominalfraser (jfr övning 1 i kapitel 2), måste vi införa en ny kategori för den fras som tillsammans med en determinator utgör en fullständig nominalfras. För denna frastyp, som exemplifieras av *tjatiga kusin* och *bok om grammatik*, används ofta kategorisymbolen N', som utläses ”N prim” eller ”N bar”. (I analogi med detta används ibland V', A' och (mera sällan) P' för fraser som utgör delkonstituenter av VP, AP respektive PP.)

(h)

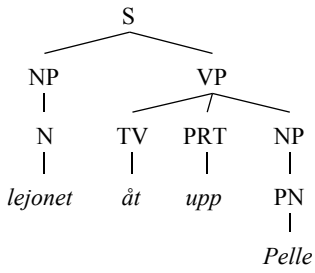


© Studentlitteratur

(i)

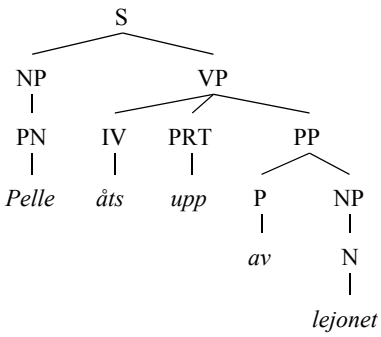


(j)



Kommentar: Kategorisymbolen PRT står för *partikel*.

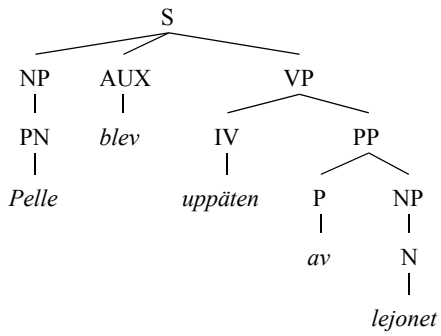
(k)

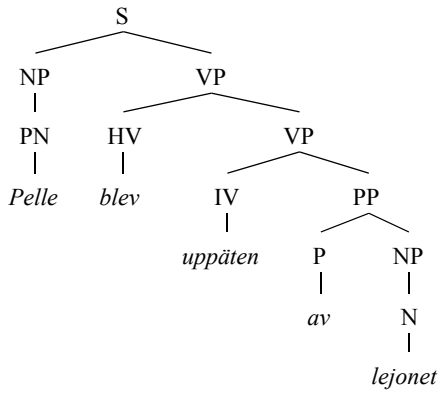


© Studentlitteratur

Bilaga A 29 January 1996 21.35:51 sida 92 av 106

(l)



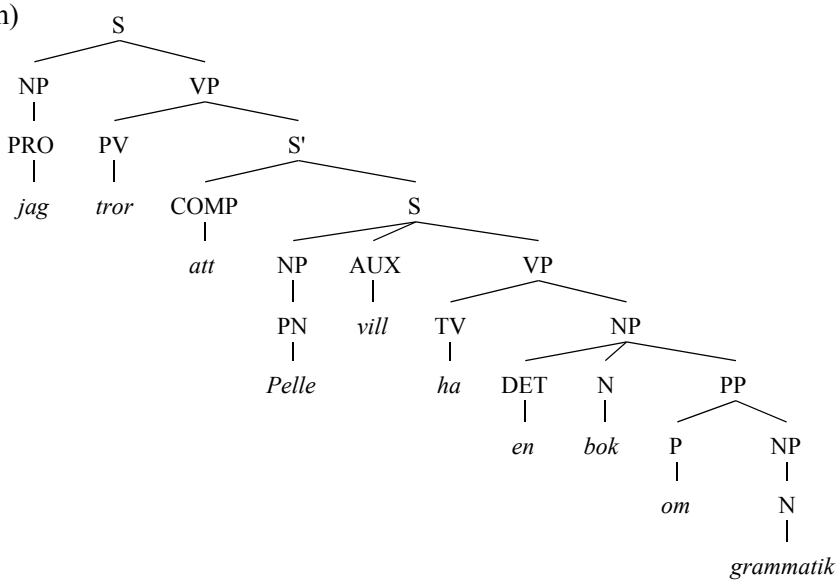


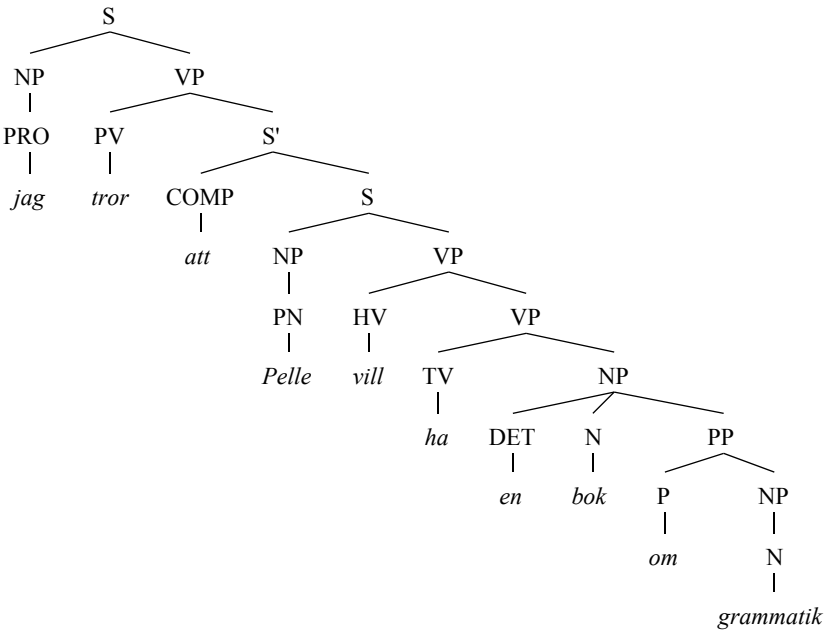
Kommentar: Som nämnts tidigare (övning 1 i kapitel 2) finns det olika uppfattningar om vilken status hjälpverb skall ges i grammatiken. Å ena sidan finns en skola som betraktar hjälpverb som en egen kategori, ofta benämnd AUX (efter engelskans *auxiliary verb*), som är skild från andra typer av verb och som förekommer som omedelbar konstituent av en sats (snarare än som huvudkonstituent i en verbfras). En annan skola anser i stället att hjälpverb i likhet med andra sorters verb fungerar som huvudkonstituent i verbfrasen. I denna analys, som återges i det andra analysträdet, har vi använt kategorisymbolen HV (för hjälpverb) eftersom kategorisymbolen AUX är intimt förknippad med det teoretiska antagandet att hjälpverb inte ingår i verbfraser.

© Studentlitteratur

Bilaga A 29 January 1996 21.35:51 sida 93 av 106

(m)



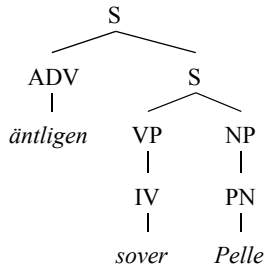


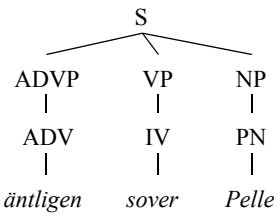
© Studentlitteratur

Bilaga A 29 January 1996 21.35:51 sida 94 av 106

Kommentar: Detta exempel introducerar tre nya kategorier. Den första är verb av typen *tror*, som tar bisatser som objekt och som vi här benämner *propositionella verb* (PV). Den andra kategorin är *bisatser*, som i den generativa traditionen brukar ges symbolen S' (”S prim” eller ”S bar”). Den tredje och sista kategorin är *underordnande konjunktioner*, som betecknas med kategorisymbolen COMP (efter engelskans *complementizer*). (Dessa tre kategorier diskuteras i kapitel 7.)

(n)



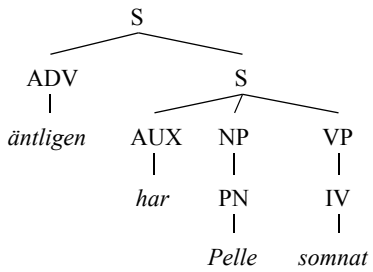


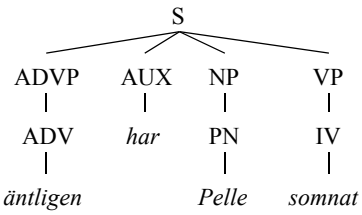
Kommentar: I denna sats förekommer adverbet *äntligen*, som vi tilldelar kategorisymbolen ADV. I analogi med övriga kategorier antar vi också att adverbet utgör huvudkonstituent i en *adverbfras* (ADVP). Vad beträffar adverbfrasens relation till resten av satsen, kan man antingen tänka sig att den utgör bestämning till en mindre sats (*sover Pelle*) eller att den utgör en omedelbar konstituent på samma nivå som verbfrasen *sover* och nominalfrasen *Pelle*. (Samma sak gäller nästföljande sats.)

© Studentlitteratur

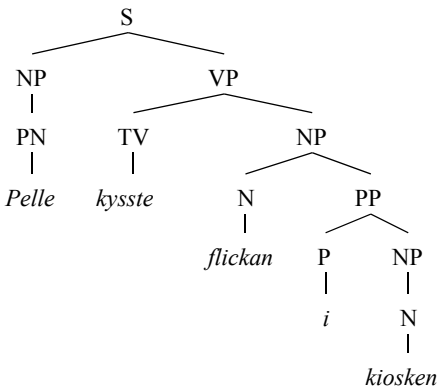
Bilaga A 29 January 1996 21.35:51 sida 95 av 106

(o)



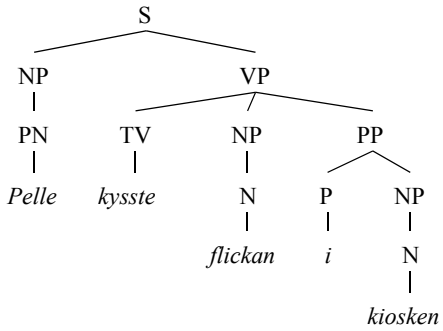


(p)

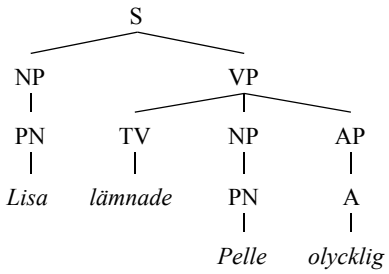


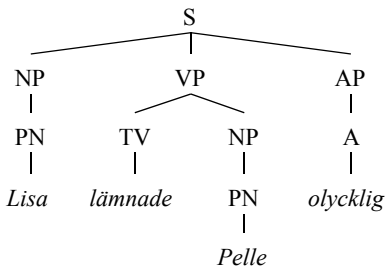
© Studentlitteratur

Bilaga A 29 January 1996 21.35:51 sida 96 av 106



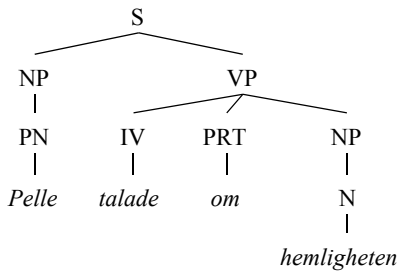
(q)

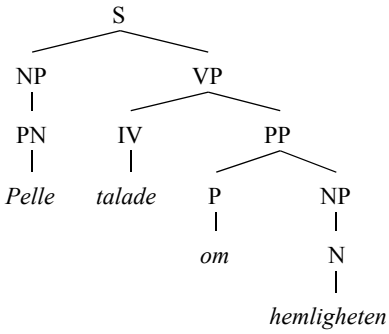




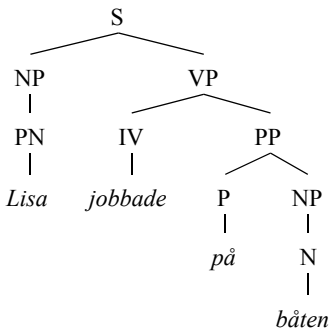
© Studentlitteratur

(r)





(s)



© Studentlitteratur

Bilaga A 29 January 1996 21.35:51 sida 98 av 106

Kapitel 4: Analysträd

Övning 1

Vi redovisar inte alla 253 relationer som råder mellan par av noder i (38), men här är i alla fall några smakprov:

Noden märkt *min* kommer före noderna märkta *bror*, *köpte*, *cykeln*, *av*, *en*, *gammal*, *vän*, N (ovanför *bror*), TV, N (ovanför *cykeln*), P, DET, (ovanför *en*), A, N (ovanför *vän*), VP, NP (ovanför *cykeln*), PP, NP (ovanför *en gammal vän*) och AP. Den kommer inte efter någon annan nod. Den dominerar inte någon annan nod. Den domineras av noderna märkta DET (ovanför *min*), NP (ovanför *min bror*) och S.

Noden märkt TV kommer före noderna märkta *cykeln*, *av*, *en*, *gammal*, *vän*, N (ovanför *cykeln*), P, DET, (ovanför *en*), A, N (ovanför *vän*), NP (ovanför *cykeln*), PP, NP (ovanför *en gammal vän*) och AP. Den kommer efter noderna märkta *min*, *bror*, DET (ovanför *min*), N (ovanför *bror*) och NP (ovanför *min bror*). Den dominerar noden märkt *köpte*. Den domineras av noderna märkta VP och S.

Noden märkt AP kommer före noderna märkta *vän* och N (ovanför *vän*). Den kommer efter noderna märkta *min*, *bror*, *köpte*, *cykeln*, *av*, *en*, DET (ovanför *min*), N (ovanför *bror*), TV, N (ovanför *cykeln*), P, DET, (ovanför *en*), NP (ovanför *min bror*) och NP (ovanför *cykeln*). Den dominerar noderna märkta *gammal* och A. Den domineras av noderna märkta NP (ovanför *en gammal vän*), PP, VP och S.

Noden märkt VP kommer inte före någon annan nod. Den kommer efter noderna märkta *min*, *bror*, DET (ovanför *min*), N (ovanför *bror*) och NP (ovanför *min bror*). Den dominerar noderna märkta *köpte*, *cykeln*, av, *en*, *gammal*, *vän*, TV, N (ovanför *cykeln*), P, DET (ovanför *en*), A, N (ovanför *vän*), NP (ovanför *cykeln*), AP, NP (ovanför *en gammal vän*) och PP. Den domineras av noden märkt S.

Noden märkt S kommer inte före eller efter någon annan nod. Den dominerar samtliga andra noder och domineras inte av någon.

© Studentlitteratur

Bilaga A 29 January 1996 21.35:51 sida 99 av 106

Övning 2

(a) Alla noder som saknar döttrar är terminala.

(b) Alla noder som har minst en dotter är icke-terminala.

(c) Alla noder som har exakt en terminal nod som dotter är preterminala.

(d) Alla noder som har minst en dotterdotter är icke-terminala och ej preterminala.

Kommentar: En typ av icke-terminala noder som inte passar in vare sig under (c) eller (d) är noder som har mer än en dotter men ingen dotterdotter (dvs. som har flera döttrar vilka alla är terminala). Sådana noder brukar normalt inte förekomma i den typ av etiketterade analysträd som används inom generativ grammatik, även om vi här inte infört några restriktioner som utesluter dem.

Övning 3

(a) Terminala symboler är symboler som etiketterar terminala noder (jfr övning 2a).

(b) Lexikala kategorisymboler är symboler som etiketterar preterminala noder (jfr övning 2c).

(c) Icke-lexikala kategorisymboler är symboler som etiketterar noder som är icke-terminala och ej preterminala (jfr övning 2d).

Kapitel 6: Ett fragment av svenska

Övning 1

För att generera (70) krävs följande regler:

```
S -> NP VP
NP -> PRO
VP -> IV
```


© Studentlitteratur

För att generera (71) krävs följande regler:

S -> NP VP
NP -> DET N
NP -> N
NP -> DET AP N
VP -> TV NP PP
AP -> A
PP -> P NP

För att generera (72) krävs följande regler:

S -> NP VP
NP -> N
NP -> DET AP N
VP -> TV NP PP
AP -> A
PP -> P NP

För att generera (73) krävs följande regler:

S -> NP VP
NP -> N
NP -> DET AP N
NP -> N PP
VP -> TV NP
AP -> A
PP -> P NP

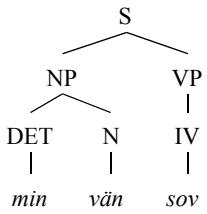
Övning 2

Samtliga ordsträngar (a–f) kan förknippas med välformade analysträd. Den sista satsen (f) kan dock endast genereras med en inadekvat struktur (se övning 3 nedan).

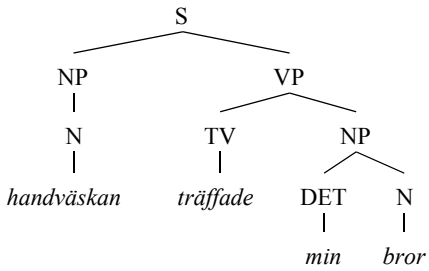
© Studentlitteratur

Övning 3

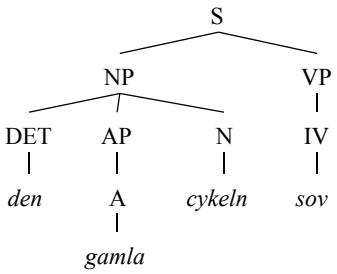
(a)



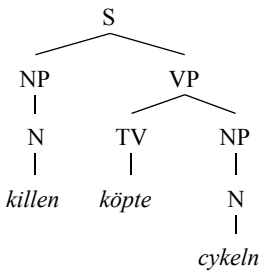
(b)



(c)



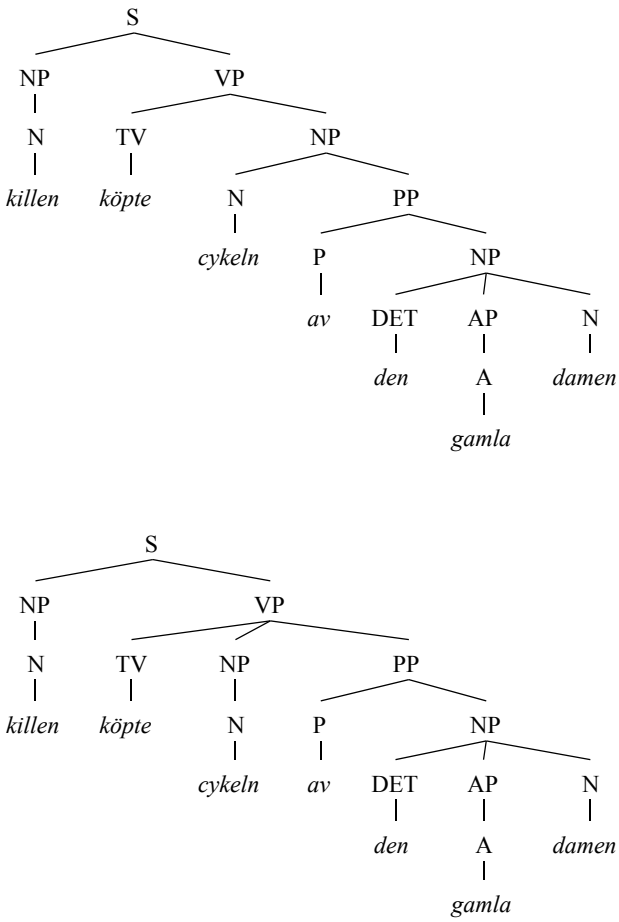
(d)



© Studentlitteratur

Bilaga A 29 January 1996 21.35:51 sida 102 av 106

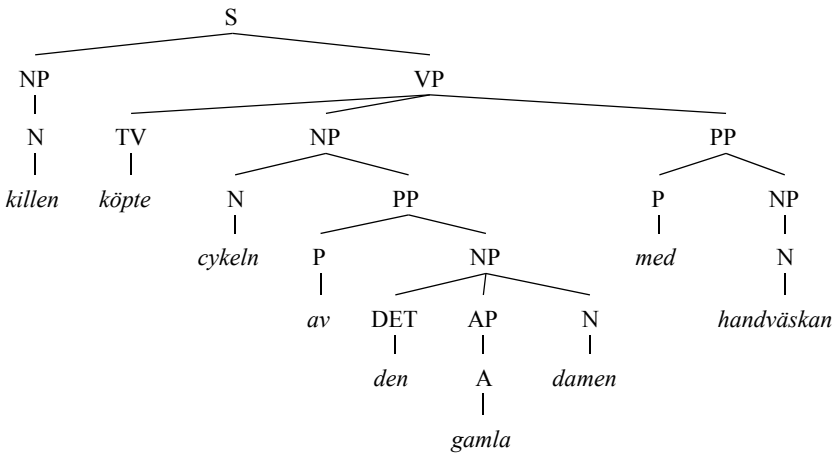
(e)



Kommentar: Strängen *killen köpte cykeln av den gamla damen* är enligt grammatiken tvetydig på samma sätt som *den gamla damen träffade killen med handväskan*. Detta stämmer inte med författarnas intuitioner.

© Studentlitteratur

(f)



Kommentar: Enligt detta analysträd, som är det enda som kan förknippas med strängen *killen köpte cykeln av den gamla damen med handväskan* enligt vår nuvarande grammatik, utgör *cykeln av den gamla damen* och *med handväskan* omedelbara konstituenter i verbfrasen. Detta stämmer inte med författarnas intuitioner enligt vilka i stället *cykeln* och *av den gamla damen med handväskan* borde vara omedelbara konstituenter i verbfrasen (jfr den andra analysen av sats e ovan). För att få denna analys måste vi utöka grammatiken med en regel som tillåter *den gamla damen med handväskan* att vara en nominalfras (vilket verkar mycket rimligt). Vi överlåter åt läsaren att formulera denna regel.

Kapitel 7: Mer om frasstrukturgrammatik

Övning 1

Reglerna kan förkortas på följande sätt:

S -> NP VP (PP)
NP -> ({DET, AP}) N (PP)
VP -> TV {NP, PP, NP PP}

© Studentlitteratur

Bilaga A 29 January 1996 21.35:51 sida 104 av 106

Övning 2

De förkortade reglerna motsvarar följande regelmängd:

```
S -> NP VP
S -> NP AP
NP -> DET N
NP -> DET AP N
NP -> DET N PP
NP -> DET AP N PP
VP -> DV NP NP
VP -> DV NP PP
VP -> DV NP NP PP
VP -> DV NP PP PP
AP -> A
AP -> A PP
PP -> P NP
PP -> P AP
```

Övning 3

Nedan anges endast de regler som behöver läggas till den ursprungliga grammatiken för att den skall generera de aktuella satserna. Analysträd för satserna gavs i samband med övning 1 i kapitel 3.

(a) Inga nya regler behövs.

(b) Inga nya regler behövs.

(c) Här behövs en ny VP-regel:

```
VP -> IV PP
```

(d) Förutom den nya VP-regeln introducerad i samband med (c) ovan krävs en regel för samordning av satser (jfr avdelning 7.2):

```
S -> S CONJ S
```

(e) Här krävs en ny regel för samordning av nominalfraser:

```
NP -> NP CONJ NP
```

© Studentlitteratur

(f) Denna sats kräver en ny VP-regel:

VP -> DV NP NP

(g) För denna sats krävs, förutom den nya VP-regeln som infördes i samband med (f) ovan, en ny NP-regel:

NP -> DET N PP

(h) Inga nya regler behövs (förutom dem som införts ovan).

(i) Inga nya regler behövs (förutom dem som införts ovan).

Övning 4

Mängden av satser är oändlig, eftersom grammatiken tillåter rekursiva strukturer, t. ex. genom regeln för samordning av satser (jfr avdelning 7.2):

S -> S CONJ S

© Studentlitteratur

Bilaga A 29 January 1996 21.35:51 sida 106 av 106

Register

A
adekvans
  deskriptiv 12
  explanatorisk 14
  observationsmässig 12
adjektiv 29, 49
adjektivfras 50
adverb 87
adverbfras 87
alternativ konstituent, se konstituent
analysträd 21

B
bisats 64

D
data 15
deskriptivt adekvat, se adekvans
determinator 29, 49
diskontinuerlig konstituent, se konstituent
ditransitivt verb, se verb
dominans 33
  omedelbar 33
dotternod, se nod

E
egennamn 80
etikett 35
explanatoriskt adekvat, se adekvans
extensionell 11

F
fonem 19
fras 22
frasstruktur, se struktur
frasstrukturgrammatik, se grammatik
frasstrukturregel, se regel

G
generativ grammatik, se grammatik
generering 11
grammatik 7, 8
  frasstruktur- 39
  generativ 7
  universell 13
grammatikalisk 9
gren 33

H
härledning 41
hierarkisk struktur, se struktur
hjälpverb, se verb
huvudkonstituent, se konstituent
hypotes 15

I
icke-lexikal kategori, se kategori
icke-terminal symbol, se symbol
intensionell 11
intransitivt verb, se verb

intuition 15

K
kategori 27
  icke-lexikal 29
  intrinsikal 28
  lexikal 29, 43
  relationell 28
  syntaktisk 27
konjunktion
  samordnande 62
  underordnande 64
konkatenering 20
konstituent 21
  alternativ 60
  diskontinuerlig 25, 69
  huvud- 31
  obligatorisk 58
  omedelbar 21
  optionell 57
konstituentstruktur, se struktur

L
lexikal insättningsregel, se regel
lexikal kategori, se kategori
lexikon 13
lexikonenhet 43
linjär struktur, se struktur

M
modernod, se nod
morfem 19

N
nod 33
  dotter- 34
  moder- 34
  syster- 34
  terminal 34
nomen 29, 49
nominalfras 29, 50

O
obligatorisk konstituent, se konstituent
observationsmässigt adekvat, se adekvans
ogrammatikalisk 9
omedelbar dominans, se dominans
omedelbar konstituent, se konstituent
optionell konstituent, se konstituent
ordklasser 29
övergenerering 68

P
partikel 75
partikelverb, se verb
precedensrelation 37
preposition 29, 50
prepositionsfras 50
preterminal symbol, se symbol
pronomen 49
propositionellt verb, se verb

R
regel 13
  frasstruktur- 40
  lexikal insättnings- 44
  rekursiv 62

transformations- 70
rekursiv regel, se regel
rekursiv struktur, se struktur
relationell kategori, se kategori
rot 34

S
samordnande konjunktion, se konjunktion
särdrag 69
särdragsstruktur 69
sats 8, 50
satsdel 28
språk 8
startsymbol, se symbol
strängderivation 40
struktur 11
fras- 21
hierarkisk 22, 24
konstituent- 21
linjär 19
rekursiv 62
strukturell flertydighet 23
substantiv, se nomen
symbol
icke-terminal 35
preterminal 36
start- 41
terminal 35
syntaktisk funktion 28
syntaktisk kategori, se kategori
syntax 16
systernod, se nod

T
teori 15
terminal nod, se nod
terminal symbol, se symbol
trädkonstruktion 42
transformationsregel, se regel
transitivt verb, se verb

U
underordnande konjunktion, se konjunktion
universell grammatik, se grammatik

V
välformad, se grammatikalisk
verb 29
ditransitivt 82
hjälp- 76
intransitivt 50
partikel- 75
propositionellt 64
transitivt 29
verbfras 29, 50

Y
yttrande 9