

..

FRED KARLSSON

Hur rekursiva är naturliga språk?

DEN ALLMÄNNA SPRÅKVETENSKAPENS grundläggande uppgift är för det första att empiriskt beskriva så många som möjligt av världens cirka 7000 språk, och för det andra att på basen av dessa grunddata formulera en generell teori för vad naturliga språk är. Av en sådan teori bör minimalt framgå vilka egenskaper som är gemensamma för alla språk och inom vilka ramar variation kan förekomma. Självfallet är en förklarande teori att föredra framom en rent deskriptiv framställning.

1. Aktuell diskussion

En av de under 2000-talet mest omdiskuterade språkliga egenskaperna är *rekursionen*. En i tidskriften *Science* år 2002 lanserad version av Noam Chomskys generativa språk teori gör rentav gällande, att rekursionen vore den enda egenskapen som är gemensam för alla språk och som inte kan härledas ur andra kognitiva förmågor. I samma anda har Derek Bickerton i en artikel i *Lingua* 2005 gjort gällande, att "[s]yntax without recursion is about as viable as Hamlet without the prince of Denmark". Men, som så ofta i språkvetenskapen, finns det också här motargument. Dan Everett lanserade i *Social Anthropology* år 2005 tesen, att det sydamerikanska aboriginalspråket Pirahã skulle sakna

(bl. a.) rekursion – och därmed falsifiera Chomskys rekursionshypotes från 2002. Den här diskussionens vågor slår höga ännu fem år senare. När man konfronteras med den självsäkerhet med vilken de extremaste ståndpunkterna förs fram kan man med fördel besinna en torr observation från 1970-talet av den i maj 2009 framlidne akademikern Nils Erik Enkvist, utländsk ledamot av Kungl. Vitterhets Historie och Antikvitets Akademien sedan 1975: ”linguists are frequently wrong but never in doubt.”

2. Vad är rekursion?

Ordet *rekursion* återgår på latinets verb *recurrere*, ’springa på nytt’. Det är fråga om upprepad tillämpning av samma process (funktion) eller konstruktion. Begreppet kom först i användning i teorin för formella språk, till exempel Gödel använde sig av rekursiva funktioner i sina berömda ofullständighetsteorem. Formellt sett är en funktion rekursiv, om den i sin definition innehåller en referens till sig själv. Mera allmänt säger man ibland, att en process är rekursiv om den upprepar sig på ett enahanda men strukturellt differentierat sätt.

En äkta rekursiv definition av en funktion innehåller två komponenter: dels ett test för att kontrollera om man nått det eftersträvade slutläget (finns inte ett sådant kommer funktionen att tillämpas för evigt), dels (ifall stopptestet inte uppfylldes) en återtillämpning av funktionen på en reducerad variant av den tidigare situationen. En skämtsam exemplifiering av detta, konstruerad av Andrew Plotkin, kan hämtas från internet: (a) om du nu vet vad rekursion är, kom ihåg det och stanna (stoppvillkoret); (b) i annat fall, leta reda på en person som befinner sig närmare Douglas Hofstadter än du är och fråga denne vad rekursion är. (Hofstadter utgav år 1979 boken *Gödel, Escher, Bach: An eternal golden braid*, där bl.a. rekursion diskuteras utförligt.) Den som vill veta vad rekursion är och i praktiken tillämpar Plotkins rekursiva funktion kommer förr eller senare antingen att stöta på någon annan än Hofstadter som kan ge det önskade svaret, eller i maximalfallet kommer det slut-

ligen att bli Hofstadter själv som ger svaret, varefter stoppvillkoret (a) uppfylls och processen avstannar.

Något mera tekniskt kan språklig rekursion definieras som i följande uppställning med användning av generativa omskrivningsregler (*rewrite rules*). Med *iteration* avses en direkt upprepning av samma process, utan den strukturella differentiering som fullödlig nästlad rekursion ger upphov till.

<i>Typ av omskrivningsregel</i> ($\alpha, \beta \neq \emptyset$)	<i>Orientering</i>	<i>Relation till iteration</i>	<i>Lingvistisk term</i>
$A \rightarrow A \alpha$	vänsterrekursion	svansrekursion = ekvivalent med iteration	vänsterförgrening = initial inbäddning
$A \rightarrow \alpha A$	högerrekursion	svansrekursion = ekvivalent med iteration	högerförgrening = final inbäddning
$A \rightarrow \alpha A \beta$	nästlad rekursion	fullödlig rekursion = icke-ekvivalent med iteration	nästling = centerinbäddning

Meningarna (1–3) nedan, alla tagna ur första ledaren i *Åbo Underrättelser* 7.5.2009, ger på satsnivå exempel på respektive vänster-, höger- och nästlad rekursion av första graden, med positionella lingvistiska termer respektive initial, final och centerinbäddning (av satser).

- (1) *Att det ska bli hårdare tag mot djurskyddsbroten* verkar klart.
- (2) *Polisen kunde inte ta kamphundarna från honom eftersom han just bytt adress.*
- (3) *Det som lagstiftningen inte rör på* är attityderna.

Initialinbäddade satser har inget material från huvudsatsen (utom eventuellt en samordnande eller underordnande konjunktion) före sig. Finalt inbäddade satser har inget material från huvudsatsen efter sig. En centerinbäddad sats har material från huvudsatsen både före (med undantag av en samordnande eller underordnande konjunktion) och efter sig.

3. Upprepad inbäddning av satser

En mening som innehåller bara en grad av satsinbäddning som i (1–3) ser förstås inte ut att vara speciellt rekursivt bildad. Ett genuint exempel med alla de tre positionella typerna av satsinbäddning är (4), där man får en föraning av rekursionens skapande kraft. Hakparenteserna anger satsernas början och slut.

- (4) [[If [when this was all over] she found the words [to tell him about it]], she wondered [if he, [who had been so sick,] would ever understand.]]

Mening (4) innehåller en dubbel initialinbäddning (*if*- och *when*-satserna), i *if*-satsen finns ytterligare en icke-finit finalinbäddning (infinitivkonstruktionen börjande med *to tell*), efter huvudsatsen förekommer ytterligare en finalinbäddning (den andra *if*-satsen), som dessutom innehåller en centerinbäddad sats, *who*-satsen.

Noam Chomsky har ända sedan år 1956 (utan empiriska argument) förfäktat tesen, att det inte finns några som helst lingvistiska restriktioner på rekursion. Inbäddningarna kan enligt honom i princip vara oändligt många och djupa både initialt (1), finalt (2) och centralt (3). Själv har jag under flera år empiriskt undersökt, om hypotesen om fri syntaktisk rekursion kan konstateras stämma (Karlsson 2007a,b, 2009, 2010). Några av resultaten ska jag presentera i det följande.

4. Upprepad initial- och finalinbäddning av satser

Vad beträffar upprepade initial satsinbäddningar finns det en mycket klar restriktion: denna operation kan utföras endast två gånger, sedan är det stopp. Med andra ord är början av mening (4) med *when*-satsen initialt inbäddad i den initialt inbäddade *if*-satsen ett exempel på den maximala påträffade komplexiteten hos initiala satsinbäddningar. En tredje initialinbäddning efter ordet *when* i *when*-satsen kommer inte i fråga, inga

sådana instanser av tredje graden har (såvitt jag vet) någonsin dokumenterats för ett enda språk.

Vad rekursiva finalinbäddningar av satser beträffar finns det ett klassiskt exempel på inbäddning ned till djupet 12 (satsinledningarna kursiverade):

- (5) This is the farmer sowing the corn, *that* kept the cock *that* crowed in the morn, *that* waked the priest all shaven and shorn, *that* married the man all tattered and torn, *that* kissed the maiden all forlorn, *that* milked the cow with the crumbled horn, *that* tossed the dog, *that* worried the cat, *that* chased the rat, *that* ate the malt, *that* lay in the house *that* Jack built.

Mening (5) är dock ett extremt unikfall. Den publicerades första gången redan år 1755 i en barnbok. Ett typiskt drag för barnboksspråket har i alla tider bl.a. varit ordlek med repetitiva strukturer, som de tolv relativsatserna efter varandra i (5), där också de talrika rimmen understryker språkleken. I talspråk är det ytterst sällsynt att påträffa rekursiva finalinbäddningar med mera än tre satser, och i skrivet språk med mera än fem. Den mest komplicerade finalinbäddningen jag påträffat i genuint skrivet språk från 1900-talet och senare är följande (6) med nio bisatser (från den brittiska korpusen ICE-GB, i något förkortad form och satsinledningarna kursiverade):

- (6) The reason ... is *that* I think *this* subject ... has importance in the lead up to ... iconoclasm *because* you know *that* this was a problem *if* that's the word *that* assailed Byzantium in the twenties *when* they convinced themselves *that* we shouldn't have portraits of Christ *shown* in human form.

Exempel av den här typen visar tveklöst, att högerförgrenande multipel satsrekursion är ett reellt språkligt fenomen.

5. *Obegränsad iteration i final position*

Formellt är rekursiva initial- och finalinbäddningar ekvivalenta med iterativa sekvenser. Med iteration avses i normalfallet upprepning av samma process utan introduktion av ny struktur. Det klaraste och vanligaste exemplet på språklig iteration är syntaktisk samordning, koordination. Samordning kan i extremfallen tillämpas hundratals gånger i samma mening, varvid mekanismen förstås är i tjänst som ett stilmedel för författaren. Författare som Broch, Joyce och Proust är välbekanta för sina långa meningar.

Till exempel i Hermann Brochs bok *Der Tod des Vergils*, utgiven 1945, finns enorma samordnade meningar, som (7), här återgiven bara som en liten excerpt. I verkligheten innehåller denna mening 1 077 ord och sprider sig ut över sex sidor, med 81 finita och 38 icke-finita satser:

- (7) ... und er, dem dieses Ich als Sinnbild, dem diese Schönheit, dem dieses Spiel, dem dieser Ablauf einverstrahlt wurde, zugestrahlt in ungetrinnbarer Notwendigkeit von den innersten und äussersten Grenzen der Welt, von den innersten und äussersten Raumesgrenzen der Nacht, so daß er dieses ganzes Geschehen in sich trug, in sich barg und zugleich davon eingeschlossen wurde, hineingehalten in den Raum der Notwendigkeit, in den Grenzraum seines Ichs, hineingehalten in den Grenzraum der Welt, in das Sinnbild ihrer Raumlosigkeit, hineingehalten in den Raum des Spieles, den Raum der überfernen Nähe, den Schönheitsraum, den Sinnbildsraum, der in jedem seiner Punkte fraglich ist und doch alle Fragen verwehrt und erstarrt, hineingehalten in alle Räume der Erstarrung, erstarrt er selber, erstickt von der Erstarrung, er fühlte, er erfaßte, daß ...

Av de finita satserna är 29 (36 %) bisatser, bl.a. 17 relativsatser, 5 *daß*-satser och 3 *als*-satser. Det maximala inbäddningsdjupet för bisatserna

är dock inte mera än 3. Meningens längd skapas framför allt av de hundratalens iterativa samordningarna, speciellt på frasnivån (och i all synnerhet gällande substantiv- och participfraser), oftast sammanfogade enbart genom juxtaposition utan samordnande konjunktioner: *einverstrahlt – zugestrahlt, von den – von den, in sich – in sich, hineingehalten* tre gånger, *Raum* i olika former åtta gånger, *erstarren* i olika former tre gånger, osv.

Iterativ samordning växer i final position 'framåt', dvs i lineär tid (i talspråket) eller högerut (i europeiska skriftspråk), utan restriktioner på antalet tillämpningscykler. Den strukturskapande rekursionen i final position, såsom i (6), är mycket mera begränsad vad beträffar tillämpningen. Exempel (6) med nio finalinbäddningar är redan unikt komplext.

6. Upprepad centerinbäddning av satser

Fullödig rekursion manifesterar sig språkligt som centerinbäddning, vilken konsumerar korttidsminnets resurser på ett mycket mera krävande sätt än initial- och finalinbäddning. Vid centerinbäddning avbryter den centerinbäddade bisatsen processeringen av den överordnade satsen, och i ett senare skede av förståelseprocessen ska lyssnaren/läsaren klara av att knyta ihop trådarna. Detta antyder, att upprepad centerinbäddning kan förväntas vara relativt sällsynt, och att upprepningarnas antal inte kan vara stort.

För flera europeiska språks del har jag lyckats belägga, att det maximala djupet för upprepad centerinbäddning är exakt 3, och till och med denna konfiguration är ytterst sällsynt, bara ett femtontal äkta exempel har dokumenterats. Meningarna (8, 9) är två sådana, från latinet och tyskan (indragningarna anger inbäddningsdjup).

- (8) Postea vos, patres conscripti,
 huic furiae, C-1
 si diutius in hac urbe, C-2
 quam delere
 cuperet, C-3
 maneret,
 vox interdiceretur,
 decrevistis ... (Cicero, *De haruspicum responsis*)

(”Efter detta beslöt Ni, Senatorer, att frånta denna galning rösträtten, om han skulle stanna längre i denna stad, som han önskade förstöra.”) Centerinbäddningarna anges med C-1, C-2 och C-3. Observeras bör, att satsen *quam delere* ’den förstöra’ inte är en centerinbäddning av grad 4; denna sats är initialt inbäddad i C-3.

- (9) Der Ritter von Malzahn,
 dem der Junker sich als einen Fremden, C-1
 der ... den seltsamen Mann, C-2
 den er mit sich führe, C-3
 in Augenschein zu nehmen wünschte,
 vorstellte,
 nötigte ihn ... (Heinrich von Kleist, *Michael Kohlhaas*)

Meningen (10) är ett svenskt exempel med dubbel centerinbäddning:

- (10) Konklaven
 som valmötet C-1
 som väljer den följande påven C-2
 heter
 kan inte ... (*Hufvudstadsbladet* 23.4.2005)

Ett framträdande drag hos alla upprepade centerinbäddningar är, att majoriteten av de inblandade bisatserna är relativa; ofta är alla relativa, som både i det svenska exemplet (10) och det tyska (9). Upprepad inbäddning av samma typ av bisatser (såsom två relativsatser, eller två *att*-satser), kallas självinbäddning (*self-embedding*). Generellt gäller, att endast satser som är efterställda bestämningar till substantiv kan vara självinbäddade; den vanligaste typen av sådana är just de relativa bisatserna och *att*-satserna. Den här restriktionen belyser på ett intressant sätt en klassisk skämtteckning av Albert Engström från årgång 1902, N:o 1, av *Strix* (se figur, reproduction KB):



Lektorns replik i Engströms teckning ser mera noggrant analyserad ut så här:

- (11) Om dig, Eriksson,
 om hvilken jag ej kunnat tänka mig något dylikt, C-1
 har jag,
 då du, C-1
 då jag till följd av iråkad
 snuva måst nysa, C-2
 brast i skratt,
 fått en högst ofördelaktig tanke.

Engströms mening (11) innehåller en självinbäddning med två *då*-satser, det vill säga en typ som står i strid med den just formulerade restriktionen att självinbäddning av identiska satser är möjlig bara om satserna är efterställda bestämmningar till substantiv. Det är uppenbarligen just brottet mot denna restriktion, som skapar det komiska intrycket, naturligtvis utöver ämnessfären och ordvalet, som förstärker effekten. En av huvudkällorna för språklig komik är normbrott.

7. Rekursion på frasnivå

Satser består av fraser, det vill säga huvudord (såsom substantiv eller prepositioner) med sina bestämmningar. Frasernas rekursiva egenskaper är inte ännu väl undersökta och därför begränsar jag mig här till två ytliga kommentarer. Iterativ initialinbäddning i fraser kan i speciella fall ske bl.a. i form av upprepade genitivattribut, som i (12). På samma sätt kan prepositionsfraser iterativt hakas efter varandra i final position, som i (13).

- (12) Sveriges nuvarande kung Carl XVI Gustaf är [*sonsons sonsons sonson*] till Karl XIV Johan. (*Hufvudstadsbladet* 5.11.2009)
 (13) Kommentarererna [*efter radioutsändningen av kungens analys av*

den rådande ordningen i Brunei] visar att ... (*Hufvudstadsbladet* 13.2.2004)

Hur det i detalj är beställt med fullödlig nästlad rekursion på frasnivå är för närvarande inte känt. Åtminstone äldre varianter av skriven tyska bjuder på tredje gradens centerinbäddning av prepositionsfraser såsom i (14), där (14a) ger hela den komplexa frasen och (14b) en närmare analys av dess inre beroendeförhållanden. Trippelinbäddning av prepositionsfraser förutsätter samtidigt tredje gradens centerinbäddning av substantivfraser, eftersom varje prepositionsfras måste innehålla en substantivfras som komplement.

(14) a. Für die nach der durch das von dem Kloster Loccum erbaute Hospitz in weiteren Kreisen bekannt gewordenen Insel Langeoog kommenden Badereisenden hat sich ...

b. Für die	PP-0; NP-0
nach der	PP-1; NP-1
durch das	PP-2; NP-2
von dem Kloster Loccum	PP-3; NP-3
erbaute Hospitz	PP-2; NP-2
in weiteren Kreisen bekannt gewordenen	
Insel Langeoog	PP-1; NP-1
kommenden Badereisenden hat sich ...	PP-0; NP-0

Det synnerligen komplexa exemplet (14) förekom i *Kaseler Zeitung* 1.5.1889. Är det en slump att den maximala gränsen för nästlad frasrekursion också förefaller att vara 3, precis som ovan noterades för upprepad centerinbäddning av satser? Denna intressanta fråga kan inte besvaras här, men uppenbart är, att en klarläggning av de kausala sambanden förutsätter en detaljerad analys av det mänskliga korttidsminnets (arbetsminnets) egenskaper.

8. Kausalförklaringar?

George Millers propagerade i sin klassiska teori från 1956 för det ”magiska talet 7 ± 2 ” som övre gräns för det antal enheter, som arbetsminnet kan lagra i verkliga situationer med tidspress. Senare minnesforskning (till exempel Nelson Cowan) har tenderat att sänka denna gräns till 4, eller 4 ± 1 . Inom denna domän ryms också den ovan konstaterade gränsen 3 för fullödlig nästlad sats- och frasrekursion.

Det är dock allt annat än klart, om satser och fraser ska betraktas som direkta processeringsenheter. Snarare är det fråga om hurdana mentala modeller mottagaren bygger upp, hur informationen blixtnabbt omkodas i större ”chunks” och över huvud taget hur gammal och ny information från den mentala analysen av fortlöpande diskurs integreras och delvis överförs till långtidsminnet.

Föredrag den 1 september 2009

LITTERATUR

- KARLSSON, FRED. 2007a. Constraints on multiple initial embedding of clauses. *International Journal of Corpus Linguistics* 12 (1): 107–118.
- KARLSSON, FRED. 2007b. Constraints on multiple center-embedding of clauses. *Journal of Linguistics* 43 (2): 365–392.
- KARLSSON, FRED. 2009. Origin and maintenance of clausal embedding complexity, i Geoffrey Sampson, David Gil & Peter Trudgill, utg., *Language Complexity as an Evolving Variable*, 192–202. Oxford: Oxford University Press.
- KARLSSON, FRED. 2010. Recursion and iteration, i Harry van der Hulst, utg., *Recursion and Human Language*, 43–67. Berlin: Mouton de Gruyter.
- KARLSSON, FRED. (2010, i tryck.) Restriktioner på meningars syntaktiska komplexitet. (Utkommer i *Språk och stil* 20.)