

Olli Blåberg  
Forskningsenheten för datalingvistik  
Helsingfors universitet

## **BÖJDA SVENSKA SUBSTANTIV I BETA**

### 1. Inledning

Svensk substantivböjning har oftast beskrivits med deklinationer. Många enskilda deklinationsmarkeringar är emellertid redundanta. Böjningsparadigmet syns ofta i vissa mönster i grundformen då genus är given. Denna artikel rapporterar om en undersökning vars syfte var att undersöka om de avvikande typerna kan räknas upp så att resultatet blir en beskrivning av den produktiva böjningstypen. Relevanta ytmönster hittas bl a i Svensk Baklängesordbok (1981). Som testmaterial har använts ett sampel på ca 2.500 substantiv ur Svenska Akademiens Ordlista (1973).

Beskrivningen är gjord i BETA, definierad och implementerad av Benny Brodda. BETA påminner om de universalmaskiner som utvecklades av matematikerna Post respektive Turing. Med universalmaskin förstås en abstrakt apparat som kan utföra alla de operationer som överhuvudtaget kan utföras algoritmiskt. Således går det att uttrycka de mest olika lingvistiska regelsystem i BETA, som inte är någon morfologisk teori i sig. Det förblir en uppgift för språkforskaren att motivera reglerna.

Regelfilen SWEPARAD i sin nuvarande omfattning består av ca 1.000 BETA regler (ca 7.000 tecken). För att tolka den har det använts en BETA implementering som är skriven i PASCAL av Kimmo Koskenniemi. Denna något mindre version får plats i en PC mikrodator. I inmatningen till SWEPARAD ges det önskade ordets grundform och genus. Eventuella sammansättningsjunkturer förutsätts vara markerade (med "="). I övrigt arbetar SWEPARAD med den ortografiska teckensträngen.

IN: EN+HÄST

UT: HÄST HÄSTS HÄSTEN HÄSTENS HÄSTAR HÄSTARS HÄSTARNA HÄSTARNAS

Många språkvetare har velat se analogier mellan datalingvistiska program och hur den mänskliga hjärnan antas processera språkliga data. Jag jämför datorer helst med pålitliga räknedosor som utför väldefinierade operationer snabbt och ofelbart.

## 2. Beskrivningsobjektet

Många lingvistiskt sett rent normativa beslut måste fattas innan regler kan skrivas. Oftast gäller det val mellan högspråkliga scheman och faktisk variation. Typisk är allomorfen -ena 'pluralis definit'. Den uppträder s g s aldrig i text, ?husena, fast suffixet är en del av (de flesta) svenskars muntliga språkbruk.

I några fall bildas böjningsparadigmet av flera stammar, t ex amerikan vs. amerikanare. Typen ansökan, ansökningar är tydligt heterotematisk. Skriftspråkets vacklan skapar ytterligare variation. T ex schejk, shejk och sjejk kan tolkas som olika "böjningsstammar" av ett lexem. Dessutom finns t ex etablerade kortformer som stan (jfr staden).

Det finns inget självklart svar på huruvida typerna arbetarn respektive arbetarena systematiskt skall ingå i paradigmet för nomen agentis suffixet -are utöver de stilistiskt neutrala arbetaren och arbetarna. Många nyare lånord saknar en stabil böjningsnorm, t ex baby (babyer / babies / bebisar) och schlager (schlagrarna / schlagersen). I princip är det omöjligt att gardera sig mot alla upptänkliga typer av framtida pluralisbildning.

I Elerts (1970, 142f.) förteckning ingår flera typer som här bedömts som främmande. Den engelska ortografins alternation y-ies (hobby, party, sherry) är jämförelsevis stabil i språket, vid sidan av den av Akademien rekommenderade typen hobbyer.

Några formella adjektiv, t ex bekant, används som vanliga substantiv. Adjektiven och verben kunde givetvis återges i BETA analogt med substantiven.

I det ortografiska alfabetet har tecknet é grafematisk status (jfr arme vs armé eller ide vs idé). Likväl finns vacklan, speciellt då formerna är entydiga, t ex armeerna och ideerna, jfr arméerna, idéerna.

### 3. Allmänt om strategin

SWEPARAD påminner till sin allmänna struktur mycket om systemet WGEN skriven av Jouko Lindstedt. WGEN genererar böjningsparadigm för finska nomen (Koskenniemi 1985). I synnerhet Hellbergs (1974, 1978) och Broddas (1979, 1982) arbeten kring svensk morfologi har också givit viktiga impulser.

Många stammar innehåller i sin form information om vilka suffix de tar. Men inte alla ord underkastar sig normalfallet, t ex omljudsorden och ord med vissa avledningssuffix. Reglerna för de avvikande orden utgör ett filter som befinner sig före normalfallets regler och släpper igenom de stammar som inte innehåller markerade mönster.

Sedan de markerade mönstren känts igen under analysens gång, blir de korrekta suffixen insatta. T ex i stammen sko hittas inga relevanta mönster och paradigmet bestäms till det ommarkerade för utrumgenus på vokal, d v s skon respektive skor. Pluralis definit skorna samt genitivformerna följer av beslutet. Ordet profil däremot har mönstret -VC(\*)VC, och böjs profilen, profiler, \*profiln, \*profilar.

I implementeringen genereras det flera ut-strängar av en enda in-sträng, vilket förutsätter förgreningsställen. Om förgreningen placeras s a s tidigt, slösas det med resurserna, vilket leder till längre exekveringstider och sämre överblick. En regel bör optimalt inte tillämpas mer än en gång.

### 4. Allomorfin i substantivböjningen

Allomorfin i böjningen är inte obetydlig. Man måste fastställa vilka allomorfer som härleds med regler och vilka som postuleras som lexikala. I SWEPARAD härleds inga ändelser från några andra med regler. Som lingvistiskt argument kan användas morfologiska "minimala" eller "subminimala par", med analogier i fonologin. Tanken har utvecklats från Karlssons (1983) begrepp kvasimorfem.

Pluralissuffixen -r, -ar och -er förekommer i s g s identisk fonologisk omgivning, i enstavningar på -ö, öar, köer samt mör. Även för suffixen -ar och -or hittas ett "minimalt par", vågar, vågor. Analogt finns nollsuffixet också i den grammatiska kon-texten utrumgenus, tekniker, \*teknikrar. T ex Svensk deskriptiv

grammatik (1981) förutsätter fem deklinationer med grundallomorferna -or, -ar, -er, -ø samt -n. I SWEPARAD inkluderas även -r bland pluralissuffixen. För singularis definit finner man "minimala par" som akademin och akademien samt knät och knäet. Lexikalt förutsätts således alla allomorferna -n, -en, -t samt -et.

En regel kunde härleda t ex -r efter vokal i stamslut från lexikalt -er: "stryk e i suffix efter vokal och morfemgräns". Antalet lexikala enheter reducerades, vilket inom vissa teorier uppfattats som målet för språkbeskrivningen. T ex mö+er skulle härledas rätt: mör, men kö+er skulle bli \*kör utan markering av stammen med någon "konsonant", deleterad senare av någon annan regel. Man kan tänka sig analoga regler för neutrum pluralis. Pluralis indefinit av segel kunde postuleras som lexikalt segel+n där suffixet -n deleterades efter konsonant (jfr häfte-n). Dock bör man inte låta typen regeln framstå som ogrammatisk.

Även en rik uppsättning systematiska allomorfer tillåter generaliseringar. Distributionen av suffixen -r, -ar och -er styrs ofta av den fonologiska kontexten, men inte alltid. Det finns en tendens enligt vilken utrum ord tar pluralissuffix på -r (-ar, -er, -r, -or) och neutrum ord suffix utan -r, men inte heller denna princip finns genomförd. Av dessa fakta dras den slutsatsen att suffixen inte bör genereras uteslutande varken på basis av den fonologiska eller grammatiska kontexten.

## 5. Grundmönstret

Det omarkerade mönstret kombinerar genusinformationen och den fonologiska informationen om ordet, t ex stavelseantalet det sista tecknets fonologiska egenskaper. I grundmönstret i SWEPARAD ingår ord med tryckaccenten på annan än sista stavelse. Tryckaccenten markeras ju inte systematiskt i ortografin.

I singularis definit får enstaviga utra suffixet -n efter vokal samt -en efter konsonant utom flerstaviga på -Vl eller -Vr, hästen, tröskeln, dollarn. Neutrum får -t efter vokal samt -et efter konsonant. Enstavingar på får vokal suffixet -et, häftet, huset, knät samt knäet.

I pluralis tilldelas alla utra på konsonant ändelsen -ar samt på vokal -r, hästar, skor. Neutra böjs med suffixen -n respektive -na efter vokaler samt nollsuffixet respektive -en

efter konsonanter, häften, häftena; tak, taken. Ytterligare har neutrum tvåstavingar på stamslutande -er det definitiva pluralis-suffixet -na, fönsterna utöver fönstren (Hellberg 1976).

Genitivsuffixet -s tillkommer efter stammar på andra tecken än s, x och z. Övriga får det ortografiska suffixet '. Undantagsvis används apostrofen i ord som i uttal slutar på ett sje-ljud, t ex Banqladesh' (jfr Banqladeshs). Banqladesh uppfattas ogärna som en genitivform. I SWEPARAD får dock inte ens de som slutar på en entydig sje-ljudssekvens någon apostrof som genitivmarkör.

I grundmönstret ingår också en morfofonologisk alternation. Senare än i första stavelse stryks e:et i sekvenserna -el, -er eller -en, ifall en suffixvokal följer omedelbart, trösklar, fönstren. Produktiviteten hos denna alternation är ett av de starkare argumenten för att både de korta och långa allomorferna för singularis definit postuleras lexikalt, d v s e:et i suffixen -n/-en respektive -t/-et varken stryks eller sätts in under derivationen. I annat fall skulle det bli avsevärda problem med regelordningen i derivationer av typen tröskeln, jfr \*trösklen.

## 6. Den ordfinala tryckaccenten

Undantagslöst (?) får utrum substantiv med final tryckaccent pluralissuffixet -er samt den definitiva singularisallomorfen -en, optionellt för stammar på vokal, jfr akademien, akademin. Ordet budget kan ta både -ar och -er då betoningen ligger på första stavelse. Om tryckaccenten placeras på andra stavelse, är endast -er möjligt. Detta visar sambandet mellan tryckaccenten och paradigmtilhörigheten. Neutra har den längre allomorfen för singularis definit optionellt för alla stammar, partiet och partit. Ingen ändelse -na finns för typen tvåstavingar på -er: klaver, klaveren, \*klavren, \*klaverna, jfr fönstren, fönster, fönsterna.

Ytterligare blockeras grundmönstrets morfofonologiska alternation: kamelen, kameler, \*kameln, \*kamler, \*kamlar; fenomenet, fenomenen, \*fenomnet, \*fenomnen. Den ordfinala tryckaccenten är ett tillräckligt villkor för avsaknaden av alternationen, men inte ett nödvändigt, siden, sidenet, \*sidnet.

Då final tryckaccent inte markeras direkt i ortografin måste man uppnå samma effekt med andra medel. T ex tvåstavingar på dubbelkonsonant markeras för detta paradigm. Mönstret -VC(\*)VCC

räknas från slutet, och de två sista konsonanterna är identiska, t ex metall, metaller. Ett relativt informativt mönster är också -VC(\*)VC, t ex profilen, \*profiln och kanyler, \*kanylar.

Några mönster är systematiskt flertydiga. Schemat -VC(\*)an räcker inte till för entydig paradigmbestämmning, afrikan men predikan. Samma gäller för -VC(\*)is: polis, poliser men tjockis, tjockisar. Lösningen är att räkna upp de ord som tillhör den numerärt mindre eller mindre produktiva gruppen. Lösningen är otillfredsställande då bägge typerna är stora och produktiva. De är olika till den fonologiska formen och sammanfaller endast i ortografin. Det finns också en typ där tryckaccenten alternerar, på första stavelse i singularis och på sista i pluralis: kansler, kanslern, \*kansleren, kanslerer, \*kanslrer.

#### 7. -ar eller -er för utrumgenus på konsonant

Relativt många substantiv böjs mot ovan presenterade regler utan några till formen reducerbara principer. Främst denna egenskap i böjningen motiverar begreppet deklination i svenskan. För flera fonologiska typer är valet mellan -ar och -er fixerat skilt för varje lexem. Det heter vild, vildar men bild, bilder och hund, hundar men kund, kunder. Ytterligare kan t ex stammar på vokal ta typiska konsonantsuffix, t ex öar och köer, eller ord på konsonant kan ta suffixet -or, t ex ros, våg, trots att suffixet i övrigt är förbehållet stamtypen utra på -a, flicka.

Framförallt -ar och -er efter konsonant i stammar med tryckaccenten på första stavelse är relevanta. De övriga typerna är små och det kan knappast hittas några regler. Några homografer skiljs åt av deklinationen, t ex bal, bank, bas, cykel, fil, form, gång, mask, regel, slav, vals o s v. Några sådana formella kriterier finns inte som kunde ge en uttömmande beskrivning av distributionen mellan -ar och -er. Dessutom finns ord där -ar och -er är i s g s fri variation, t ex checker och checkar.

Grova mönster kan dock stipuleras. Dessa verkar vara mera eller mindre godtyckliga, men nyttiga tendenser i lexikonet. Vad som synkront ligger bakom vilket suffix som fixeras för vilket lexem torde sakna ett enkelt svar. Linell (1972, 32f.) noterar att det ofta är ordets innehåll (och härstamning) som motiverar dess böjning. Abstrakta och vissa inlånade substantiv tenderar

att ha -er böjning. Bl a paret balar eller baler ger en intuition om detta. En beskrivning är alltså endast en återgivning av en lista på undantag, inte en förklaring av fenomenet.

Några tecken är så gott som irrelevanta då de av fonotaktiska eller ortografiskt-historiska skäl sällan förekommer i svenska ordslut, typiskt c, h och w. Uppenbarligen av språkhistoriska skäl verkar -er dominera i stammar på vissa dentalkonsonanter, t ex -d (kod, skillnad), -s (vas, klass) och -t (helhet, stat).

Det förutsätts undantagslistor och submönster. Det finns t ex ord på -d som tar -ar, sked, värld, tråd m m. Tillvägagångssättet är analogt i normalfallet: orden på -g tilldelas pluralis -ar, dag. Men bl a de som slutar på -rg får suffixet -er, färg, utom undantagen, korg. Dels skall undantagen till normalfallet räknas upp, d v s sak, saker trots normalfallets burk, burkar, dels undantagen till undantagslistorna, hytt, hytter men hatt, hattar. Systemet fungerar inom ramen för det vokabulär som utgjort utgångspunkten. Men då det inte blivit klart vad som egentligen styr valet av antingen -ar eller -er, finns inga garantier för att alla nya ord kommer att gå enligt samma mönster.

#### 8. Typerna flicka och pojke

De språkhistoriskt jämförbara typerna flicka och pojke beter sig olika i datorbeskrivningar, jfr Hellberg (1978, 18). Typen flicka kan hanteras med ett fåtal triviala regler. Varje utrum stam på -a bildar singularisformerna med a:et kvar i stammen och pluralisformerna utan a, flickan, flickor. Enda komplikationen kommer från det faktum att de intermediära representationerna för typen kamera inte deltar i den morfofonologiska alternationen, således kameror, \*kamror. De egentliga undantagen till reglerna för stamslutande -a är få, t ex historia, kollega.

Typen pojke kan inte ges någon uttömmande beskrivning. Ordfinalt -e i singularis kan signalera tre olika paradigm: pojke, pojkar; linje, linjer (normalfallet) eller idé, idéer. Då är idiosynkratiska (avlednings-) suffix, t ex -are, -ande samt -ge, garage, redan borträknade. Enda alternativet är att gå igenom alla mönster som kan utnyttjas. Det kan tyckas att resultatet saknar elegans, men det fungerar.

## 9. Några komplikationer

Singularis definit är oftast regelbunden. Dock saknar t ex substantiven med -an efter verbrötter definita singularissuffix, t ex ansökan, \*ansökanen. Utra med final tryckaccent på -n får ofta ingen markör för definit singularis i tal (Thorell 1977, 31f.). Formen telefon (även enstaviga, scen) kan tolkas som antingen indefinit eller definit. Pragmatiskt kan man motivera några paradigm som saknar markörer för definit singularis, bland dem månadsnamnen, t ex januari. Undvikande av homonymi verkar i vissa fall vara den bästa tillgängliga motiveringen, t ex bit är inte singularis definit av bi, men knät är en form av knä.

Ett ords genus syns i singularis definit. De ord som alternativt har båda definita suffixen har två genus, t ex paraply. Valet av pluralissuffix predestineras däremot inte av genus. Visserligen finns det en stark tendens till det grundmönster som diskuterades ovan, men ofta är det antingen ett genomskinligt avledningssuffix eller rentav fonologiska mönster som är starkare indicier för val av pluralisändelserna. Jag ser ingen orsak att tala om "genusbytare" (jfr Teleman 1969, 172; Linell 1972, 33).

Bland utrumgenus är det i synnerhet avledda substantiv med -are, -ande och -(ik)er som tar oväntade suffix, arbetare, handlande, tekniker, inte \*arbetarar, \*handlander eller \*teknikrar.

Bland neutrum är iögonenfallande t ex flerstaviga neutra på betonat -i som oftast har pluralis -er, t ex kansli. Dock böjs t ex staffli i pluralis definit stafflien eller stafflierna. Även orden på -um är speciella: antingen enligt typen faktum, fakta eller museum, useer. Analysen görs mera problematisk av tendensen att böja även dessa enligt grundmönstret. Formen fakta skall normativt ersättas med grundmönstrets pluralis indefinit faktum. Ibland träffas även typen museumet vid sidan av useet.

I några synkront genomskinliga sammansättningar böjs efterleden annorlunda än som fristående stam. Som exempel kan nämnas jungfru, jungfrur enligt grundmönstret, inte \*jungfruar. Efterleden kan också böjas mera markerat än den fristående formen: bokstav, inte \*bokstavar utan bokstäver.

Massorden verkar ofta inte ha ens potentiella pluralissuffix. T ex substantivet mat kan heta varken \*matar, \*mater eller \*mator i pluralis. Detta visar att det inte enbart är



formen som betingar böjningen hos alla ord utan det är i innehållet man hittar motiveringen till denna morfologiska egenhet.

I vissa ortografiska ställningar dubbleras respektive för-enklas ordfinala -m och -n: program, programmet, \*programet; mun, munnen, \*munen. I den mån det finns s k intern böjning i svenskan, täcks den inte av SWEPARAD, t ex prins-gemål, prinsen-gemålen samt kronprins-regent, kronprinsen-regenten.

#### 10. Avvikande paradig

I svenskan återfinns tendensen att det ofta är högfrekventa, centrala begrepp i kulturen som har ett avvikande böjningsmönster. Ord som man, moder, son, finger, öra, fot, huvud, sommar, morgon, afton, djävul, bok, grej, bryter alla mot grundmönstret. Några egentliga formella eller semantiska kriterier verkar inte finnas. Givetvis implicerar tendensen inte att alla benämningar på "centrala begrepp" måste ha avvikande böjning. T ex kroppsdelsbetecknande ansikte, hår, arm, knä, tå, släktskapsbetecknande syster, fast och tidsbetecknande dag, dygn, vinter, vår, höst m m böjs enligt grundmönstret.

Karaktäristiskt är singularisformerna oftast regelbundna. Men även avvikande singularis definit förekommer, t ex sommaren, djävulen, papperet, himmelen utöver sommarn, djävuln, pappret, himmeln, som går enligt grundmönstret (jfr dollarn, konsuln, numret, muskeln). I pluralis kan avvikelserna däremot vara relativt stora och idiosynkratiska. Exempelvis paradigmet man, män, männen, även mannar, är ett unikum.

Det verkar finnas delgrupper också bland de avvikande orden. Det finns t ex många omljudsord på -and, and, brand, hand, rand, strand, tand, land (neutrum !). Få talare skulle dock vara villiga att acceptera \*sänder som pluralis av massordet sand.

Enda möjligheten att återge dessa avvikande paradig är att ge en förteckning över dem på ett eller annat sätt. Det skulle emellertid innebära en stor belastning för regelapparaten och därmed också för den exekverande implementeringen om man tvingades räkna upp varje enskild böjd form. Därför har det visat sig vara förnuftigt att markera vissa lexikala ingångar med "morfofonem" som möjliggör endast en lexikal representation för t ex omljudsorden av typen stad, städer.

## 11. Sammanfattning

Modern svensk substantivböjning motiveras av många faktorer. Dels finns grammatisk genus, häst-en, hus-et. Dels väljs suffixen på basis av stammens fonologiska form, häst-ar, sko-r, idé-er. Dels finns fonologiskt omotiverade paradig, bal-ar, bal-er, och paradig där godtycklig fonologiskt betingad paradigmatillhörighet är starkare än genus, kansli-et, kansli-er, \*kansli-n. De sistnämnda typerna motiverar egenskapen deklination hos substantiven. Det förekommer också fonologiskt motiverad alternation i stammarna, muskel, muskler, som saknas i regel då de fonologiska villkoren inte är uppfyllda, kamel, kameler. Ytterligare upprätthålls ett stort antal idiosynkratiska paradig, bok, böcker, där ingen motivation står att finna i stammens form.

Den svenska böjningen är ett språkligt delsystem som måste motstå tryck, av vilka t ex främmande språk och inhemsk variation inte är de minst betydelsefulla. Således finns optionalitet också i böjningen, betingad av många olika faktorer. Denna variation accentuerar språkbeskrivningens allmänna normativa karaktär.

Erfarenheterna från SWEPARAD visar att det inte finns någon enkel lösning på datoriserad svensk paradigmgenerering. Men det är faktiskt möjligt att konstruera ett tillfredsställande system som kan utvidgas utan svårighet och som fungerar väl på en mikro-dator. BETA har återigen visat sig vara ett utmärkt lingvistiskt verktyg för morfologisk datorbeskrivning.

Datalingvistiken tvingar fram explicita beskrivningar. Den syn på svensk substantivböjning som tillämpats här avviker på flera punkter från gängse uppfattning, men kan likväl försvaras. Å andra sidan är det klart att flera olika beskrivningar kan alla fylla sin funktion. T ex pedagogiska målsättningar behöver inte nödvändigtvis sammanfalla med datalingvistiska syften.

## Litteratur

- Brodda, B. (1979), Något om de svenska ordens fonotax och morfotax: iakttagelser med utgångspunkt med experiment med automatisk analys. Papers from the Institute for Linguistics, University of Stockholm 38.
- - - (1982), Yttre kriterier för igenkänning av sammansättningar. Svenskans beskrivning 13. Meddelanden från institutionen för nordiska språk och nordisk litteratur vid Helsingfors universitet, serie B nr 6. 102-114.
- Deskriptiv svensk grammatik (1981), Red. Holm, B. & Nylund, E. Trelleborg. 10:e tryckningen.
- Elert, C. (1970), Ljud och ord i svenskan. Uppsala.
- Hellberg, S. (1974), Graphonomic Rules in Phonology. Nordistica Gothoburgensia 7.
- - - (1976), Bestämd form pluralis. Nysvenska studier 55-56:227-250.
  - - - (1978), The Morphology of Present-Day Swedish. Stockholm.
- Karlsson, F. (1983), Suomen kielen äänne- ja muotorakenne. Juva.
- Koskenniemi, K. (1985). A system for generating Finnish inflected word-forms. I: Computational Morphosyntax. Report on Research 1981-1984. Red. Karlsson, F. Publications of the Department of General Linguistics, University of Helsinki, no. 13. 63-80.
- Linell, P. (1972), Remarks on Swedish Morphology. Reports from Uppsala University Department of Linguistics nr 1.
- Svensk Baklängesordbok (1981), Allén, S. & Eeg-Olofsson, M. & Gavare, R. & Sjögreen, C.. Nacka.
- Svenska Akademiens Ordlista över svenska språket (1973), 10:e upplagan. Stockholm.
- Teleman, U. (1969), Böjningssuffixens form i nusvenskan. Arkiv för nordisk filologi. 84:e bandet. Lund.
- Thorell, O. (1977), Svensk grammatik. 2:a upplagan. Stockholm.