

# Vad jag i min verksamhet som språkvårdare skulle vilja att datorlingvistiken bidrog med

Margareta Westman  
Stockholm

Ska det här vara ett utbyte så måste jag börja med att förklara vad vi språkvårdare håller på med. Inte minst eftersom många människor, inklusive lingvister, ofta har en lite skev uppfattning om vad språkvårdare egentligen gör.

En av våra uppgifter är att tala om för folk hur de ska tala eller skriva, och det gör vi direkt, till exempel per telefon, eller indirekt genom handledningar, ordböcker, språkspalter och radioprogram. Däremot gör vi det aldrig utan att vara ombedda, tillfrågade. Utöver ren rådgivning verkar vi också för att hålla i gång en allmän diskussion om språket, för att höja medvetenheten.

Man kan undra hur det kommer sig att folk över huvud taget frågar hur något ska heta eller vad ett ord betyder. Alla som kan språket kan ju språket – det är ett axiom, tror jag, i språkvetenskapen.

Ändå är det ju så att väldigt mycket av vårt språkkunnande är inlärt på ett mer systematiskt sätt än det rent spontana samtalspråket är. Riktigt fullärd blir man väl heller aldrig ens i sitt eget språk. Detta hänger i sin tur ihop med att språket inte är statiskt. Levande språk förändras ständigt.

Den språkform vi språkvårdare yttrar oss om är standardspråket, alltså det språk vi alla har lärt oss som en övernorm som används för att skriva och också tala i mera offentliga eller stora sammanhang.

Hur kan vi då upplysa folk om hur det ligger till? Vi måste grunda våra råd på bruket, vi kan inte förlita oss enbart på fåtöljlingvistisk intuition. Vi måste faktiskt samtidigt både misstro vår intuition – och utnyttja den.

För att kunna ge bra råd behöver vi kunskap på tre nivåer av språket. Vi behöver känna till

- 1) bruket, hur folk gör, dvs. faktiskt talar och skriver,

- 2) uppfattningen, hur folk tror att de gör, dvs. de allmänna åsikterna om hurdant språkbruket är,
- 3) idealet, hur folk anser att man bör göra, dvs. människors idealbild av språket.

De här tre olika aspekterna är viktiga alla tre. De ingår alla på olika sätt i språkkunnandet och i många fall måste vi undersöka alla nivåerna.

### Först något om bruket

Den självklara lösningen vore att undersöka förhållandena i en textkorpus och räkna. Så enkelt är det dessvärre inte. Man kan konstatera frekvenser av det ena eller det andra. Men hur är det vid närmare skärskådande? Är de observerade skillnaderna helt fritt varierande eller finns det subtila betydelskillnader i olika textkontexter? Eller skillnader i olika sociala kontexter? Frågan är alltså: Vad är egentligen **samma** fall?

Problemet kan illustreras med ett exempel. Jag tar upp uttrycket *vare sig – eller* som är en av de vanliga käpphästarna för språkriktighetsivrare, samtidigt som det är principiellt intressant eftersom det gäller frågan om förhållandet mellan negation och nekande innebörd.

Det här uttrycket används ofta utan något *inte*, trots att en negation av hävd har ansetts nödvändig för att den riktiga innebörden ska förmedlas.

"Äsch", säger ni, "ett enskilt uttryck, det är ingen konst att fånga upp via någorlunda stora korpusar. Det fixar vi lätt."

Och visst kan man få ledning av fynd i stora korpusar, men fynden måste också analyseras. I det här fallet t ex, tycks det finnas en skillnad mellan konstruktioner då *vare sig – eller* föregås av ett finit verb som kan ta en negation intill sig och konstruktioner där *vare sig – eller* i sin helhet är infogat i en inledande nominalfras:

*Språket är inte så lätt att förstå, vare sig i teori eller i praxis.*

respektive

*(Inte?) Vare sig i teori eller praxis är språket lätt att förstå.*

I det första fallet kan inte negationen slopas; däremot händer det ofta i fall av den andra typen.

När vi så med hjälp av många textexempel tycker oss ha funnit en möjlig distinktion behöver den prövas. För att värdera exempel som de här behöver vi tillgång till bedömningar av andra än oss själva. Med det menar jag inte lingvisters tyckanden, utan mer systematiskt samlade åsikter. Och därmed är vi inne på nivå 2 och 3, uppfattningarna och idealen.

Det här första exemplet är ändå rätt enkelt, det är ju inte så svårt att söka ett enskilt uttryck. Det finns knepigare fall där bruket tycks hålla på att förskjutas. Vi kan ta frågan om *-a* eller *-e* på adjektivet i bestämd form, alltså, sådant som:

*den amerikanske/amerikanska författaren Susan Faludi*  
*den ryske/ryska källan meddelar*  
*universitetets nye/nya datasnille*

Traditionell språkvård avvisar kategoriskt *e*-form när nominalfrasen syftar på en kvinna, när huvudordet inte är ett personord eller när huvudordet är ett substantiv i neutrum.

För att få fram ordentligt med exempel på sådant här behövs satslösta, dvs parsade, eller på något sätt preparerade korpusar. Även så är det rätt besvärligt att analysera de olika fallen, för det är många faktorer som tycks samspela: referensen, huvudordets betydelse och morfologi, adjektivets betydelsetyp, hela frasens genretillhörighet och dessutom gamla regionala språkskillnader.

Därtill eller kanske på grund av allt detta tror jag att det finns genuina åsiktsskillnader om hur man får eller bör göra, vilket vi också behöver kartlägga.

Rent syntaktiska problem är naturligtvis inte lättare att konstatera bruket av. Jag har t.ex. svårt att föreställa mig hur jag skulle kunna få fram relevant material över hur subjeksregeln tillämpas vid infinitivuttryck genom att göra datorsökningar.

En sak är helt dock klar: Vi behöver kunna utnyttja stora textkorpusar av skilda slag, från olika genrer och tider. Stora, väldefinierade och verkligen spridda över genrer och tid på så vis att man kan få jämförbarhet. Spridning över genrer är inte minst viktigt för vårt stora projekt med en konstruktionsordbok. I den ska vi visa vilka typer av bestämmningar som olika ord måste ha respektive kan ta och vilka fraser, mer eller mindre fasta, de kan ingå i.

Jag undrar också om man inte borde försöka få med det man kan kalla vuxnas folkliga skrivande, alltså sådant som klubbtidningar (kan gälla

husdjur – sport – frimärken ), annonsblad, föreningsprotokoll. På så vis skulle vi kunna få inblick i hur texter ser ut när ingen central kontroll varit verksam.

Detta om bruket.

## **Om folks uppfattningar**

Det nästa vi behöver är alltså kunskap om folks uppfattningar och ideal. Hur samlar och registrerar man data om vad folk anser att de gör? Detta är faktiskt viktigt. Mycket inläring av språk vilar på det, inklusive vår egen inläring av skriftspråket och främmande språk.

På Språknämnden hade vi en gång en idé om att upprätta en panel av språkkunnigt folk eller folk som talar och skriver i offentliga sammanhang och registrera deras uppfattningar (jfr inledningen i *American Heritage Dictionary*). De här personerna skulle få sig underställda olika språkproblem som är på tapeten, både gamla inkörda och nya, och få bedöma vad som går an och vad som inte går an. Det gäller alltså någon sorts regelbundet återkommande språkmentometer som kanske kunde skötas via datorer. Så kunde man med jämna mellanrum få en översikt där åsiktstrycket på olika punkter registrerats. Helst skulle man förstås vilja stämma av dessa personers egen praxis mot deras tro om sin praxis och även deras uttryckta ideal.

En annan källa till kunskap om uppfattningar är litteraturen om språk. Här kommer våra planer på en databas som kunskapsbank in. I den ska läggas in för det första referenser till språkvetenskaplig litteratur, handböcker och språkspalter. Där ska också in vår egen excerpering av nya ord och uttryck, konstruktioner osv. Banken ska tjäna oss på två sätt, dels i den dagliga rådgivningen, dels i vår grundforskning som behövs som underlag för utgivning av ordböcker och handböcker.

Om och när vi någonsin får råd med den utrustning som behövs kommer vi också att behöva bistånd av datalingsvister för att lägga upp det hela på det sätt som är klyftigast för våra behov, och för andras, ty ytterligare ett syfte med basen är att göra våra samlingar tillgängliga för andra forskare.

Ett behov här är bättre scanningsteknik, dels för att föra över våra kort i basen, dels för att föra in nya excerpter.

Vårt riktigt stora problem just nu är emellertid åtkomligheten, eller den bristande åtkomligheten hos existerande textbaser. Vi vet ännu inte om vi – utan att behöva betala väldigt årliga avgifter som vi inte har råd med –

kan få komma in på universitetsdatabaserna. Hittills har de varit slutna för oss. Numera kommer allt fler korpusar på CD, och det skulle kunna vara en lösning. En vanlig invändning mot den lösningen är att man då inte är inne i de senaste uppdateringarna. Det är emellertid en klen invändning om alternativet är att man inte är inne alls.

## Om nya ord

Jag vill gärna också ta upp ytterligare en forskningsdel av vår verksamhet. Språknämnden ska enligt sina stadgar följa svenska språkets utveckling i tal och skrift – som den anspråklösa uppgiften är formulerad.

Vi registrerar alltså nyheter i språket, bland annat nya ord. Det är ju på det området som saker händer snabbast och tydligast.

Nya ord tillkommer på olika sätt, som gör det olika svårt att komma åt dem genom någon sorts automatisk excerpering. Det kan gälla:

1. Helt nya ordformer som *deska* (arbeta med desktop), *krockkudde* (i bilar), *roligan* (om snäll, lugn fotbollsentusiast), *tjugolapp* (om tjugokronorssedeln),
2. Nya fraser
  - a) med nya ord i som *varken bu eller bä* (varken det ena eller det andra),
  - b) utan nya ord i som *gröna räkenskaper* (om bokföring där man räknar in miljökostnader), *fiska danskt* (som går så till att man sätter sig en bit upp från stranden med matsäcken och låter den yngste i laget vakta metspöna och ropa om det händer något)
3. Gamla ord i ny betydelse som *tjuga* (om tjugokronorssedeln), *golv* (om undre gräns),
4. Nya bildliga användningar som *logga ut* ( för 'dö'), *ta fram* (för 'utarbeta')

Den första gruppen borde vara lättast att datorfinna men den blir väl överbelastad av alla tillfälliga sammansättningar – som i och för sig kan vara intressanta. De övriga typerna måste vara mycket besvärliga att söka automatiskt. Eller?

Men – än en gång – vi behöver verkligen korpusar och konkordanser – tillgängliga och lättanvända.

### **Till slut en fundering**

Om vi skulle få tillgång till perfekta korpusar med spridning över tider och genrer, finns det då en risk att vi alla kommer att medverka till att låsa språkutvecklingen vid status quo? Normeringskraften hos t.ex. ordböcker är ju stark. Riskerar vi rundgång på ett sätt som aldrig hittills har inträffat?

Jag tror att risken eller chansen är liten – men det är verkligen viktigt att normering och råd grundar sig på bredast och djupast möjliga insikter i språklivet och inte på diverse idiosynkratiska föreställningar. Och för detta behöver vi hjälp av datorlingvister.