

Språkforska på nätet

Hur vanligt är det att erfarna skribenter börjar meningar med *och*? Vad är vanligast i dag, *för de som* eller *för dem som*? Finns ordet *funktionalitet*, och vad står det i så fall för? Hur använde Strindberg svordomar i sina brev och romaner? Kan man hitta slangord som *slagga*, *jäkta* och *käka* hos äldre svenska författare? Ordböcker och handböcker ger oss inte alltid vägledning i språkfrågor. Men på nätet kan du i dag själv snabbt få fram svar på alla tänkbara typer av språkfrågor. Här följer en presentation av några av de bästa hjälpmedlen för den som vill göra egna små språkundersökningar.

Denna handledning kommer att uppdateras med jämna mellanrum (se webbsidan www.spraknamnden.se). Senast uppdaterad 17 oktober 2001. Författare [Ola Karlsson](#), Svenska språknämnden, juni 2001.

Sökmetoder

Sök i textarkiv på ord och fraser

I textdatabasen [Språkbanken](#), liksom i de kommersiella arkiven [Presstext](#) och [Mediearkivet](#), kan du på webben enkelt och snabbt undersöka hur ord och uttryck faktiskt används. Här hittar du både dagstidningar, kulturtidskrifter och tekniktidningar; i Språkbanken även författningstext och mängder av skönlitteratur. Du kan göra egna undersökningar och t.ex. se vilket år ett visst ord börjar dyka upp eller om ett ord eller en viss stavningsvariant har blivit vanligare de senaste åren. En fördel är förstås att exemplen ingår i autentiska texter så att man också får ett sammanhang. Liknande textdatabaser finns även för andra språk, t.ex. [Bank of English](#) för engelska.

Vill man veta om ett uttryck är accepterat eller inte är det bra att veta att ett utbrett språkbruk, i synnerhet i språkmedvetna tidningar som Dagens Nyheter, Göteborgs-Posten och Svenska Dagbladet, ofta även accepteras av språkvården. Men inte alltid - ibland betraktas ett utbrett språkbruk fortfarande som felaktigt, och i de fallen kan det förstås vara värdefullt att se hur vanligt felet är.

Gör grammatiska sökningar i textarkiv

Många språkfenomen handlar om konstruktionstyper snarare än om specifika uttryck, varför man ibland behöver söka på grammatiska kategorier, som substantiv eller infinita verb. Det kan man göra i [Parole](#), en textsamling där alla ord försetts med information om ordklass och böjning. Vill man veta hur vanligt det är med futuralt *kommer* utan *att* kan man då t.ex. söka på *kommer* följt av verb i infinitiv. (Sökningen i Parole är dessvärre än så länge ganska svårhanterlig.)

Att använda Språkbanken och Parole (DN, SvD, GP m.m.) är gratis. Presstext (DN, Expressen, GT, Veckans affärer m.fl.) och Mediarkivet (Aftonbladet, SvD, Arbetet m.fl.) kan t.ex. skolor prenumerera på till en låg kostnad. För närmare information om hur du använder arkiven, se instruktionerna på respektive webbplats.

Sök i söktjänster

Vill man titta på språkbruket på nätet - tidningsspråket är ju långt ifrån alltid representativt för språkbrukarna i gemen - kan man göra frekvensberäkningar i de stora söktjänsterna, t.ex. på [Altavista](#) eller [Google](#). Särskilt lättanvänd och snabb är Google. Där kan man dessutom söka separat i [diskussionsforum](#), som ofta ger ett annat resultat än annan webbtext. Sökbara diskussionsgrupper hittar man även på [Yahoo](#) (under [Avancerat sök](#)). Se respektive söktjänst för sökinstruktioner. Är ett uttryck du söker i en text inte markerat, hittar du det lätt med din webbläsares sökfunktion ([sök/find](#) under redigeramenyn).

Svenska Akademiens ordbok på nätet

Den mest omfattande svenska ordboken är utan jämförelse [Svenska Akademiens ordbok](#). De 31 hittills utgivna banden är både dyra, tunga och svåra att slå i, så gratisversionen på nätet har stora fördelar. Här får man en detaljerad beskrivning av hur olika ord, fraser och former använts sedan 1500-talet, dessutom med uppgifter om ordens historia och första textbelägg. Det tar lite tid att lära sig använda ordboken, men med lite läsarbete ser man t.ex. snart att *innan* använts som preposition redan på medeltiden (i stället för *före*) och att talspråksformerna *eran* och *erat* har funnits länge i svenskt skriftspråk.

Andra elektroniska lexikon

Utöver SAOB finns det mängder av elektroniska lexikon med innehåll som kan vara svårt att hitta i tryckt form. Två bra samlingar av ordlistor finner man hos [Ordboken.nu](#) samt på [Svenska språkbyråns](#) webbplats. Med hjälp av dessa källor kan man få svar på många frågor: Vad heter *attachment* på svenska? Vad står *BSE* (galna ko-sjukan) för? Heter det *enolog* eller *önolog*? Hur används ordet *benchmark* i engelskan?

- På [Skoldatanätet](#) finns t.ex. Lexinseriens svensk-kroatiska och svensk-finska lexikon och ett svensk-dansk-norskt lexikon.
- Helengelska lexikon hittar man bl.a. på [Skoldatanätet](#), på [Onelook](#) och på [Yourdictionary](#).
- Termordlistor för olika fackområden är viktiga. På webben finns bl.a. [Svenska datatermgruppens](#) och [Svenska biotermgruppens](#) ordlistor, liksom EU:s gigantiska termdatabas [Eurodicautom](#) med svensk terminologi på en rad fackområden.

Övrigt nätmaterial

På nätet finns hur mycket material som helst som kan användas för språkliga efterforskningar.

- Ordböcker behöver t.ex. ibland kompletteras med sökningar i encyklopediska verk som [Nationalencyklopedin](#) (kräver prenumeration), [Nordisk familjebok](#) (begränsat innehåll) eller [Britannica online](#), t.ex. för att hitta ett geografiskt namn eller för att få mer information om fenomenet bakom ett ord.
- Söker man mer spännande textmaterial är det inte svårt att hitta. Hos bl.a. Projekt [Runeberg](#) finns stora mängder äldre nordisk skönlitteratur i elektronisk form. Här kan man t.ex. göra mer detaljerade sökningar efter citat och fraser i olika verk eller undersöka hur olika äldre svenska författare brukat vissa ord och uttryck. Hos [Eboken.nu](#) och [The on-line books page](#) kan du gratis ladda ner mängder av modern litteratur.
- På nätet finns dessutom mängder av gratis språkhjälp: i [Svenska Dagbladets](#) arkiv kan du söka i språkspalter (sök på [Språkspalten](#)) och på [Språknämndens](#) webbplats kan man få svar på vanliga och aktuella språkfrågor under [Frågor och svar](#).

Söktips

För att få mer exakta resultat bör man snäva in sin sökning. Vill man söka efter en specifik fras sätter man den inom **citattecken**, t.ex. **"lägga rabarber på"**. Söker man efter två eller fler separata ord kan man använda **variablerna AND, OR och NOT** eller deras motsvarigheter i den aktuella tjänsten; ibland används t.ex. plus (+), komma (,) eller **och** för AND, minus (-) eller **inte** för NOT och **eller** för OR. Skriver man **hund AND katt** ger det texter med båda dessa ord. Söksträngen **hund NOT katt** söker texter med ordet *hund* som inte innehåller ordet *katt*. **Hund OR katt** tar fram texter med antingen ordet *katt* eller *hund* eller båda. Alternativt anges inom parentes: skriver man **hund (katt OR gås)** ger det texter med orden *hund* samt antingen *katt* eller *gås*. Ska orden förekomma nära varandra i texten anges det ofta med **NEAR** (eller **vid**). I vissa tjänster kan man också ange om orden ska stå i samma mening eller stycke och hur många ord som högst får skilja de båda sökorden åt. Observera att det bara är det andra ordet som ska föregås av en sökvaribel, inte det första (skriv t.ex. **hund -katt**, inte **+hund -katt**).

För att hitta både sammansättningar och olika böjningsformer av ett ord använder man s.k. **trunkering** och sätter en asterisk (*) efter eller före grundformen. Skriver man **prestanda*** får man alltså *prestandan* och *prestandatopp* medan ***prestanda** ger *topprestanda* och *maxprestanda*. Frågetecken (?) eller punkt (.) används ibland som variabel för ett utelämnat tecken: med söksträngen **Fran??n** kan man hitta olika stavningsvarianter av *Fransén* (med *s* eller *z*, *e* eller *é*).

Ibland talas det i textdatabaser om **ordförekomster** (eller tokens) och **ordtyper** (eller types). Ordtyper är alla de enskilda ord och ordformer som finns i databasen, t.ex. *pojken* och *pojkar*, medan ordförekomsten är det totala antalet ord eller träffar på ett visst ord (eller ordform): 20 träffar på förekomsten *pojken* och 25 på förekomsten *pojkar* ger sammanlagt 45 träffar på ordet *pojke*.

Exempel

Seoul eller Söul?

Stavning av namn framgår sällan av ordböcker. Undrar man om Sydkoreas huvudstad stavas *Seoul* (den engelska formen, tillika den Nationalencyklopedin anger) eller *Söul* kan man t.ex. göra en sökning i [Presstext](#). Logga in och klicka på **Starta sökning**. Börja med att skriva in **Seoul** i sökordsfältet och klicka på **Sök**. Det ger 83 exempel på denna form, av vilka många kommer från Expressen. Skriver man in **Söul** i stället, visas 2 214 belägg. *Söul* är alltså mer än 25 gånger vanligare i svensk tidningstext, vilket indikerar att det är den helt dominerande, sedan länge vedertagna svenska formen.

Ska eller skall?

Hur har bruket av *ska* och *skall* sett ut över tiden? Vi undersöker saken i [Språkbanken](#) och skriver med valet **Sök i: frekvenstabell** in söksträngarna **ska :n** och **skall :n**, där **:n** ger en jämförbar frekvens för de olika materialen (alla konkordanser räknas som om de bestod av en miljon ord). Värdena gäller antal träffar, vilket anges i sidhuvudet.

Det ger följande siffror för *ska*: P65 (tidningstext 1965) 339, P76 (tidningstext 1976) 1 128, P97 (tidningstext 1997) 1 840, moderna Bonniersromaner ca 1 700 träffar, och för *skall*: P65 2 024, P76 2 093, P97 1 138, moderna Bonniersromaner ca 500 träffar.

Trots att *skall* dominerar stort i 60-talets tidningsprosa (i materialet ser vi att Svenska Dagbladet fortfarande huvudsakligen skriver *skall*), uppvisar Äldre romaner (Almqvist, Bergman, Lagerlöf m.fl.) faktiskt lika många *ska* som *skall*. Söker man på *ska* i [SAOB](#) (sök på *skola* i sökfältet under **Uppslagsord** och sedan *skola.v3* eller på *ska* under **Fritextsökning**) hittar man snabbt skriftliga belägg på *ska* från 1600-talet. Tvärtemot mångas föreställning har *ska* alltså funnits mycket länge samtidigt som *skall* lever i all välmåga.

(I Svensk författningssamling runt 1980 finns inte ett enda *ska*; vill man undersöka bruket av *ska* och *skall* i modernt myndighetsspråk kan man göra det i fritextsökningar i samlingar av protokoll, propositioner etc. direkt på [riksdagens](#) webbsidor.)

MP3-aktig eller MP3-mässig ljudkvalitet?

Suffixen *-mässig* och *-aktig* är mycket produktiva och används bl.a. för att karakterisera saker, ofta genom liknelser: *amatörmässig*, *sjuttiotalsmässig*, *konkursmässig*, *rödaktig*, *flickaktig*, *slampaktig*. De har alltså ungefär samma innebörd. Men hur gamla är de? Har de flesta orden med *-mässig* funnits i flera hundra år eller är flertalet nya? Är *-mässig* eller *-aktig* det vanligaste suffixet, förr och i dag? Vid en sökning i [SAOB](#) under **Fritextsökning** eller **Kombinerad sökning** (med kronologisk inställning under **Sortering** och 200 under **Antal träffar**) ser vi att ***mässig** ger färre träffar och att beläggen härrör från början av 1700-talet och framåt, medan ***aktig** ger betydligt fler träffar, av vilka många härrör ända från 1500- och 1600-talet. Söker man dessutom på **aktig** under **Uppslagsord** får man fram den artikel som handlar om suffixet *-aktig*, och där framgår att det funnits redan i fornsvenskan.

Tittar vi sedan i [Språkbanken](#) får vi följande värden: *-mässig* ger i P97 214, i P65 26 (materialet i P97 är ca 12 gånger större än det i P65) och i Äldre svenska romaner bara 8 träffar; *-aktig* ger i P97 367, i P65 48 och i Äldre svenska romaner 201 träffar.

I [Presstext](#) ger ***aktig** 9 401 och ***mässig** 4 670 träffar. För att se om något hänt mellan 1997 och i dag kan vi i Presstext gå till **Avancerad sökning** och göra en tidsbegränsad sökning, förslagsvis en för 1997 (**970101-971231**) och en för 2000 (**000101-001231**). Det ger 1 003 *-aktig* och 495 *-mässig* 1997 samt 1 211 *-aktig* och 565 *-mässig* 2000. Något ökat bruk av *-mässig* mellan 1997 och 2000 föreligger alltså inte.

Orden med *mässig-* tycks ha slagit igenom först under 1900-talet och har under senare tid blivit nästan lika vanliga som orden med *aktig-*.

Börja göra eller börja att göra?

Några hjälpverbliknande verb, som *börja*, *försöka*, *hinna*, *behöva*, *tänka* m.fl., får ofta i likhet med de äkta hjälpverben inte *att* framför det infinitiva verbet: *får göra*, *försöker (att) göra*. Tar dessa verb i regel *att* eller inte? Kan man se en tendens för bruket med eller utan *att*?

För att få en lättöverskådlig samling belägg på konstruktion med *att* kan man förslagsvis börja med att söka i [Parole](#). Söksträngen **"försök. |börja." [msd="CIS"] [msd="V@NOAS"]** innebär att programmet söker på olika former av *börja* och *försöka* (punkten anger att flera tecken kan följa, t.ex. *började*, *försökt* etc.) följt av *att* och ett verb i infinitiv. Under **Taggtabell** hittar man färdiga söksträngar för ett antal ord och ordklassformer. Den aktuella strängen ger 225 träffar (55 *börja*, 169 *försöka*). Konstruktion med *att* är alltså inte sällsynt, även om en sökning utan *att* ger oerhört många fler träffar (6 731). Eftersom det infinitiva verbet inte alltid följer direkt efter *att* kan man skriva in markörer för ytterligare ord: **"försök. |börja." [msd="CIS"] [] [msd="V@NOAS"]** ger ytterligare 24 fall, **"försök. |börja." [msd="CIS"] [] [] [msd="V@NOAS"]** ytterligare 3.

Vill vi sedan söka tendenser i bruket går vi lämpligen till [Språkbanken](#). Eftersom vi inte kan göra en grammatisk sökning som i Parole, får vi välja

en vanlig fras (ger väldigt få träffar) eller manuellt räkna alla *börja/försöka* respektive *börja/försöka att*.

I P97 hittar vi uppskattningsvis ett tusental fall av *börja* och *försöka* + infinitiv men bara 5 *börja att* och 23 *försöka att* + infinitiv. I P65 hittar vi 13 *börja* och ca 50 *försöka* + infinitiv men inget *börja att* och bara 1 *försöka att* + infinitiv. I [SAOB](#), som ju huvudsakligen innehåller ett äldre språkmateriale, får vi t.ex. 13 *börja göra* och 1 *börja att göra*. Någon ökning av konstruktion med *att* tycks alltså inte föreligga.

E-post, e-mail, mail eller mejl?

Ingen skriver väl *e-post*, utom möjligen några ordentliga dagstidningar? undrar någon. Vi går till [Google](#) och skriver in de olika orden i tur och ordning och klickar på [Sök svenska sidor](#). För att inte *mail* ska få träff även i *e-mail*, får man skriva [mail -e-mail](#) (dvs. hitta *mail* men ta bort alla *e-mail*). Resultatet visar att *e-post* faktiskt är klart vanligare (ca 585 000 belägg) än *mejl* (23 200), *e-mail* (168 000) och *mail* (155 000) i svenska webbtexter.

Ett helt annat resultat får man vid en diskussionsgruppssökning på [Google](#). Klicka på länken [Advanced Groups Search](#). Välj sedan [svenska](#) under språkmenyn och skriv in respektive ord i sökfältet (för att inte *mail* ska få träff även i *e-mail*, fyll i [e-mail](#) i fältet [utan dessa ord](#)). Det ger 74 000 *mail* och 33 500 *e-mail*, men bara 12 000 *e-post* och ynka 1 300 *mejl*. Språkbruket i de informella diskussionsgrupperna skiljer sig alltså markant från det i andra webbtexter.

Är site samma sak som web site?

Site används i svenskan för eng. *web site*, trots att *site* egentligen bara betyder 'plats'. Men hur är det, används kanske *site* som kortform för *web site* även i engelskan? För att ta reda på det går vi först till [Bank of English](#). Man kan få maximalt 40 träffar i demoversionen, men söksträngen [site+8web](#) (hittar *site* i texter som också handlar om webben, med maximalt 8 ord mellan *site* och *web*) ger bara 3 träffar, där *site* dessutom står mycket nära ordet *web*:

The site is [httpwww.le-web.fr](#); the most useful site on the World Wide Web; your own site on the World Wide Web sounds exciting.

Vi kan av resultatet sluta oss till att *site* som kortform för *web site* är sällsynt i engelskan, och det tycks dessutom förekomma mest i ett givet sammanhang. Söker vi däremot på ["web site"](#) i t.ex. Altavista får vi över 5 miljoner träffar.

En policy, flera vadå?

Någon undrar hur man böjer *policy* i plural men ni hittar inget entydigt svar i ordböckerna. I stället går ni till [Språknämndens](#) frågor och svar, klickar på [p](#) och sedan på [policy](#) och ser där att det kan böjas *policyer*.

Den nya ekonomin, vad är det?

I undervisningen ska ni diskutera uttrycket *den nya ekonomin* men ingen vet riktigt vad det betyder och ni hittar ingen definition i svenska lexikon. Ni

söker då på [One look dictionaries](#), skriver in "new economy" i sökfältet och hittar underlag till en egen definition i tre engelska ekonomiordlistor.

Hur ser meningar ut som Strindberg börjar med *och* och *men*?

Många tror att man inte får börja meningar med *och* och *men* (det får man!). Tittar man efter upptäcker man emellertid att det är ett vanligt grepp hos många erkända författare. I Språkbanken går det inte att söka på meningsbörjan. Därför går vi till projekt [Runeberg](#).

Klicka på länken [Search](#) och fyll i fältet **Title:** på nästa sida i namnet på en roman som finns med i Runebergsamlingarna, t.ex. [Röda rummet](#). Klicka på sökrutan ([Search](#)) och när sökresultatet sedan kommer upp på länken [Röda rummet](#). För att kunna söka i hela texten och inte bara i ett kapitel i taget, rulla ner till slutet på innehållsförteckningssidan och klicka på länken som visar texten som html-fil ([HTML file](#)). Nu kan du i din webbläsares vanliga sökfunktion ([sök/find](#) under redigeramenyn) skriva in söksträngarna [. Och](#) respektive [. Men](#). När du fått upp det första exemplet, gå tillbaka till [sök/find](#) och klicka igen för att komma till nästa exempel.

Som synes börjar Strindberg mycket ofta meningar med *och* och *men*. Särskilt många fall finns i dialogtexterna, dvs. i återgivet tal. Det ser man även om man gör samma sökning i Strindbergs dramer. Ett par exempel:

"Han tvättade sig i ansiktet och om händerna och kammade sig; en polis skulle ha tagit honom för en mördare som höll på att utplåna spåren efter sin gärning. **Men** under allt detta tänkte han - klart, bestämt och redigt. **Och** när han fått rummet och sig själv i ordning hade han fattat ett beslut, vilket han verkligen länge hade förberett och som nu skulle sättas i verket."

"**Men** solen stod över Liljeholmen och sköt hela kvastar av strålar mot öster; de gingo genom rökarne från Bergsund, de ilade fram över Riddarfjärden, klättrade upp till korset på Riddarholmskyrkan, kastade sig över till Tyskans branta tak, lekte med vimplarne på Skeppsbrobåtarna, illuminerade i fönstren på Stora Sjtullen, eklärerade Lidingöskogarne och tonade bort i ett rosenfärgat moln, långt, långt ut i fjärran, där havet ligger. **Och** därifrån kom vinden ..."

"På nedra botten till vänster om porten var en kristlig bokhandel. Falk stannade och läste på de i fönstret utlagda böckernas titlar! Det var det gamla vanliga: indiskreta frågor, närgångna tillvitelser, förnärmande förtroligheter, allt så väl och så länge känt. **Men** vad som mera ådrog sig hans uppmärksamhet var de många illustrerade tidskrifterna, vilka med sina stora engelska träsnitt lågo utbredda för att locka folk."

Gosse, grabb, pojke, kille

Hur vanliga är orden *gosse*, *grabb*, *pojke* och *kille* - förr och i dag? Svaret på en sådan fråga kan man få på [Språkbanken](#). För att få ett grovt intryck av frekvenserna för orden jämför man förslagsvis materialen press 65, press 76 och press 97. Skriv in ett ord i taget i sökfältet och klicka på **Sök i: frekvenstabell**. Lägg till **:n** efter varje sökord för att få en jämförbar frekvens för de olika materialen (alla konkordanser räknas som om de bestod av en miljon ord). Sedan kan man enkelt ställa upp en tabell över de olika ordens frekvens i materialen.

	P65	P76	P97	Totalt
gosse	10	3	2	15
grabb	2	5	4	11
pojke	22	32	25	79
kille	4	19	27	50

Här ser vi t.ex. att *gosse* var avsevärt vanligare 1965 än i dag, att *kille* blivit betydligt vanligare de senaste decennierna även i tidningsprosan och att *pojke* och *grabb* (som är ovanligt) förekommer ungefär lika mycket i dag som förr.

Observera att alla sifferuppgifter och adresser i handledningen refererar till sökningar gjorda i mars 2001