


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Jagten på det gode citat - Om vanskelighederne ved at finde egnede ordbogseksempler i et korpus	
Forfatter:	Henrik Lorentzen	
Kilde:	Nordiska Studier i Lexikografi 5, 2001, s. 202-216 Rapport från Konferens om lexikografi i Norden, Göteborg 27.-29. maj 1999	
URL:	http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive	

© Nordisk forening for leksikografi

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Søgbarhed

Artiklerne i de ældre Nordiske studier i leksikografi (1-5) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

Henrik Lorentzen

Jagten på det gode citat

Om vanskelighederne ved at finde egnede ordbogseksempler i et korpus

The paper deals with the difficulties involved in selecting good dictionary examples from a large text corpus, especially if all examples are to be taken directly from the corpus without modifications. The author gives examples of the problems that the lexicographer has to cope with when he is put face to face with authentic language, such as length, disturbing elements, difficult words, errors, and lack of naturalness. Thus, the emergence of text corpora has not made the lexicographer's work easier, but the result will hopefully be a better dictionary with better examples.

Indledning

Denne artikel handler om sprogbrugseksempler i en stor monolingval ordbog og i særdeleshed om de vanskeligheder der er forbundet med at udvælge autentiske eksempler i form af citater fra et større, elektronisk søgbart tekstkorpus. Umiddelbart virker det jo tillokkende at have en stor tekstmængde til rådighed bestående af "rigtigt" sprog, sætninger som folk faktisk har skrevet eller sagt. Nu skulle man endelig få mulighed for let at finde eksempler på sproget som det bruges ude i verden, modsat sådan som ordbogsredaktøren forestiller sig sproget. Virkeligheden er imidlertid lidt anderledes. Ganske vist er alle de forekomster man møder, autentisk, faktisk forekommet sprog, men det er ikke ensbetydende med at de egner sig til at være citater i en ordbog. Ud af flere hundrede forekomster kan det ofte være svært at finde bare et citategnet belæg, bl.a. fordi de autentiske belæg netop forekommer i en sammenhæng og derfor har svært ved at fungere som løsrevne ordbogseksempler. I artiklen vil jeg give eksempler på de forhold der kan diskvalificere et citat, og i den sammenhæng se på citatets funktion i ordbogsartiklen.

Mine erfaringer og mit eksempelmateriale stammer fra mit arbejde som redaktør på Den Danske Ordbog, et projekt under Det Danske Sprog- og Litteraturselskab. Arbejdet med denne nye danske modersmålsordbog har nu været i gang i knap 8 år og forventes afsluttet i 2002, hvor ordbogen begynder at udkomme i 6 bind. Den Danske Ordbog skal efterfølge *Ordbog over det Danske Sprog* (ODS), som beskriver dansk indtil ca. 1955. Ordbogen bygger primært på et elektronisk tekstkorpus der dækker perioden 1983–92. Ud over at ligge til grund for beskrivelsen af ordenes betydninger skal korpus også levere autentiske eksempler, citater, til illustration af ordenes betydning og brug (se evt. mere om Den Danske Ordbog på Internettet: www.dsl.dk/ddo.html)

Når man ser bort fra ODS, er det noget nyt at en dansk almenordbog benytter autentiske citater som eksempler. Den hidtidige praksis synes at have været konstruerede, redaktionelle eksempler, evt. i form af bearbejdede autentiske eksempler. I litteraturen diskuteres det indgående om man skal foretrække redaktionelle eller autentiske eksempler. Tilhængerne af redaktionelle eksempler holder på at ordbogsredaktøren med sin store viden om sproget er den bedste til kort og klart at eksemplificere det pågældende opslagsords brug, uden at blande alt for mange forstyrrende elementer ind i billedet. Fortalerne for autentiske eksempler mener derimod at de konstruerede eksempler netop virker konstruerede, abstrakte og kunstige i forhold til sætninger der påviseligt er forekommet i faktisk sprogbrug og kan bringe en ny friskhed som de konstruerede og måske stereotype eksempler ikke har (Bergenholtz 1994). Ikke mindst britiske leksikografer (især folkene bag Cobuild) har gjort sig til talsmænd for kun at bruge autentiske eksempler, nu hvor disse er så let tilgængelige via elektroniske korpora.

Den Danske Ordbog, som altså også bygger på et korpus, indeholder også (stort set) kun autentiske eksempler i form af citater. Vi lægger os hermed i forlængelse af de eksisterende store modersmålsordbøger i Skandinavien (ODS, *Svenska Akademiens ordbok*, *Norsk Riksmålsordbok*), som favoriserer de autentiske eksempler.

Hvad er et godt citat?

I det følgende vil jeg beskæftige mig med hvordan et godt citat ser ud, og hvad der gør et citat mindre egnet til at stå i en ordbog.

Før man kan afgøre hvad et godt citat er, må man først overveje hvad citatets funktion er. Hvis citatet blot skal dokumentere at ordet findes i sproget på et bestemt tidspunkt, behøver man ikke stille særlig store krav til det. En forekomst som opslagsord i en ordbog, et leksikon eller en ordliste kan fx være tilstrækkeligt. Skal citatet derimod supplere eller illustrere de allerede givne oplysninger om ordet, dvs. bidrage yderligere til forståelsen af ordet og vise det i naturlig sammenhæng med andre ord, viser det sig hurtigt at langt de fleste autentiske ytringer ikke er velegnede.

Sven-Göran Malmgren har i en artikel i *LexicoNordica* (Malmgren 1994) opstillet fire kategorier af ordbogseksempler, inddelt efter hvilken type information der gives i dem. Disse fire kategorier er:

- (1) Semantiske eksempler.
- (2) Syntaktiske eksempler.
- (3) Konnotationelle eksempler.
- (4) Encyklopædiske eksempler.

Semantiske eksempler understøtter og supplerer betydningsbeskrivelsen ved at vise ordet sammen med betydningsmæssigt beslægtede ord, fx synonyme eller antonymer, eller sammen med ord som det typisk optræder sammen med (kollokationer), fx for et substantivs vedkommende de adjektiver der typisk beskriver det, og for et verbs vedkommende typiske subjekter og typiske objekter.

Syntaktiske eksempler viser de almindelige syntaktiske forbindelser som ordet indgår i, fx et verbum konstrueret med direkte objekt, eller forbundet med præpositioner og i givet fald hvilke. Ofte vil et givet eksempel både illustrere semantiske og syntaktiske forhold, fx hvis et transitivt verbum optræder med et typisk objekt: *Hun spiller klaver.*

Konnotationelle eksempler antyder typiske associationer, følelser og situationer som opslagsordet er knyttet til eller giver anledning til, fx ved ordet *fødselsdag*, hvor et godt eksempel vil kunne handle om fest og gaver.

Encyklopædiske eksempler giver faktisk, ikke-sproglig viden om det begreb som ordet refererer til, fx statistiske oplysninger om forekomsten af en sygdom.

Jeg ved ikke om denne rækkefølge er tilfældig fra Malmgrens side, men det stemmer med min egen opfattelse at et godt citat primært bør illustrere opslagsordets naturlige semantik og syntaks, mens konnotationer og encyklopædiske oplysninger kommer i anden række.

Når korpus kun består af autentisk sprog, burde det jo være let at finde eksempler på naturlig sprogbrug. I praksis viser det sig dog at være ret vanskeligt at udvælge gode citater. Jeg har forsøgt at inddele problemerne i en række kategorier som i høj grad modsvarer de krav jeg selv stiller til et godt citat, nemlig:

- (1) Længde.
- (2) Forstyrrende elementer.
- (3) Svære ord.
- (4) Manglende naturlighed med hensyn til sprog eller situation.
- (5) Fejl.
- (6) Tendentøse eller anstødelige udsagn.
- (7) Manglende underholdningsværdi.

Jeg vil i det følgende give konkrete eksempler på nogle af disse problemer, idet der dog ofte er tale om en kombination af flere forhold.

Længde

Det er en god gammel leksikografisk grundregel at man skal spare på pladsen. Eksempler er ingen undtagelse til denne regel. De fleste redaktionelle ordbogseksempler er korte, mens auten-

tiske eksempler har en tendens til at blive ganske lange, især hvis man (som vi på Den Danske Ordbog) tilstræber at alle citater indeholder et finit verbum. Man kan dog ofte forkorte ved at udelade passager der ikke er direkte relevante for opslagsordet, og erstatte disse med prikker. Vi tillader os også at resumere passager og erstatte dem med en kort omskrivning i firkantet parentes, men man kan hurtigt komme i konflikt med kravet om autenticitet. Hvor meget må man gøre ved citatet og stadigvæk påberåbe sig at det er autentisk? Og endelig hvad betyder det for læsbarheden at citaterne indeholder underlige tegn og gerninger i form af prikker og parenteser? For at dæmme op for dette har vi på Den Danske Ordbog ikke tilladt at kombinere disse forkortelsesmetoder, men alligevel ser vi ofte at ønsket om at holde længden nede giver en høj frekvens af prikker og parenteser. Eks. 1 viser hvordan man kan gøre vold på et belæg for at forkorte det, men uden at resultatet er specielt vellykket. Opslagsordet er *klarsyn*:

- (1) Problemet er blot at finde konkrete psykiske fænomener og dermed personer, som kan tillægges paranormale evner. Det vil sige evner, der ikke kan forklares ud fra kendte naturvidenskabelige principper. For eksempel tankelæsning (telepati), forudviden (prækognition) og klarsyn (clairvoyance) (Michael Leslie Clemmensen: *Er der mere mellem himmel og jord?*, 1992)

Ved hjælp af parenteser og prikker har man forsøgt at redigere denne lange smøre ned til:

- (1a) [paranormale evner] vil sige evner, der ikke kan forklares ud fra kendte naturvidenskabelige principper. For eksempel .. klarsyn (clairvoyance)

Citatet er ganske vist blevet meget kortere – til gengæld er der både brugt firkantede parenteser og prikker, foruden de runde parenteser som citatet i forvejen indeholder. Et sådant belæg er ikke muligt at redigere på en fair måde og egner sig ikke særlig til ordbogsbrug, også fordi det ikke primært handler om opslagsordet *klarsyn*, men blot nævner det i en opremsning sammen med andre paranormale evner.

Eks. 2 er både kortere, lettere tilgængeligt og frit for redaktionelle indgreb:

- (2) Guds profet Samuel var en sådan "seer", der i sit klarsyn kunne se, hvor bortløbne dyr befandt sig (Den Nye Dialog, 5/1992)

Forstyrrende elementer

Autentiske eksempler indeholder ofte mange oplysninger og elementer som er relevante i den konkrete kommunikationssituation hvor sætningen er blevet brugt. Det kan fx være navne på personer og steder, årstal, andre tal, og det kan være deiktiske udtryk og tekstmarkører af typen *her*, *så*, *desuden*, *derfor*, *således*. Lige så naturlige og nødvendige som den type oplysninger er i faktisk sprogbrug, lige så unaturlige og uvedkommende kan de forekomme når citaterne bruges, helt løsrevet fra deres kontekst, i en ordbog hvor der skal illustreres almene forhold vedrørende ordenes betydning og brug. Her et eksempel vedr. ordet *ordensregel*:

- (3) Ø C K og K O K er imidlertid indforstået med at udarbejde ordensregler, hvorefter medlemmerne skal parkere på parkeringspladsen nord for Svanholmskolen (Frederikssund Avis, 1983)

I princippet er citatet godt fordi det illustrerer hvad ordensregler kan omfatte (her regler for parkering), men det fungerer ikke så godt uden for kontekst fordi man ikke kan regne ud hvad Ø C K og K O K står for, og måske endda bliver lidt forvirret af det (der er tale om to sportsklubber).

Et bedre citat ville være det følgende, som stammer fra en lejekontrakt:

- (4) Lejeren skal have et eksemplar af ordensreglerne og skal følge disse regler (Labor Lejekontrakt, 1983)

Vi får – også her – sneget en væsentlig encyklopædisk oplysning ind, nemlig at der findes ordensregler i boligejendomme, til gengæld indeholder citatet ingen forstyrrende, uigennemskuelige

forkortelser, og endelig er det kun halvt så langt. I eks. 5 ser vi et andet tilfælde hvor elementer peger ud af teksten på en forvirrende måde (belæg på ordet *ulykkelig*):

- (5) Derfor var man langt ringere stillet end i dag, hvis man var så ulykkelig at blive ramt af sygdom (Sven Frøkjær-Jensen: *Segl og scepter – danskernes virkelighed*, 1991)

Uden for kontekst er det selvfølgelig umuligt at vide hvad *derfor* henviser til, og citatet er af den grund ikke velegnet i ordbogen.

Svære ord

Vi tilstræber på Den Danske Ordbog at det ordforråd og den syntaks der bruges i definitionerne, er så ligetil som muligt. Lidt anderledes kan det forholde sig med citaterne. Her er det jo sproget i "fri dressur" som vi afspejler, og derfor er det usandsynligt at alle ord og udtryk fra citaterne vil figurere som opslagsord i ordbogen. Ikke desto mindre bør citaterne ikke være for svære at tilegne sig da de ellers ikke vil fremme forståelsen af definitionerne, og i værste fald måske endda blokere for forståelsen. Ordet pesticid er – af gode grunde – blevet meget udbredt i samfundsdebatten, men det følgende belæg på ordet er for indviklet (og for langt) til virkelig at gøre gavn:

- (6) ved at overføre det toxinkodende gen, håber det store amerikanske frø- og kemikaliefirma Monsanto at have et mikrobielt pesticid til bekæmpelse af jordboende sommerfuglelarver (Morten Kvistgaard et al.: *Bioteknologi og miljø*, 1987)

Mange brugere vil nok have problemer med ordene *toxinkodende*, *mikrobiel* og *jordboende* sådan at opmærksomheden fjernes fra opslagsordet selv. Formentlig har man større glæde af følgende citat.

- (7) Kommissionen anbefaler, at forbruget af pesticider i landbruget formindskes (Statens Information: *Status*, 1986)

som ud over at være enklere og kortere også nævner landbruget, der jo er kendt som storforbruger af pesticider.

Manglende naturlighed med hensyn til sprog eller situation

Som redaktør vil man ofte være optaget af at finde et citat hvor flest mulige af Malmgrens fire kategorier kommer med (semantik, syntaks, konnotation og encyklopædi). Især kan man være fristet af citater hvor der gives encyklopædiske oplysninger, men her risikerer man at fjerne sig for langt fra den naturlige sprogbrug. Det er nemlig ret sjældent at man i almindelige samtale-situationer beskæftiger sig med hvordan diverse begreber helt præcist defineres, eller hvilken familie en bestemt blomst henregnes til. I følgende eksempel beskæftiger man sig fx med den præcise meteorologiske definition af fænomenet "hvid jul":

- (8) En hvid jul betyder, at mere end 90% af Danmark skal være dækket af mindst 1/2 cm sne den 24. december om eftermiddagen (AKB-beboeren, 4/1998)

Der er ingen tvivl om man her får megen encyklopædisk information, men til gengæld er det ikke et eksempel på naturlig brug af udtrykket *hvid jul*. I vores korpus beskæftiger man sig ganske vist i nogle få tilfælde med statistikken for hvid jul i Danmark i 1900-tallet (ca. hvert tiende år i øvrigt!), men mere udbredt er det at drømme om eller håbe på en *hvid jul*, så et mere karakteristisk citat ville få fat i udtrykkets konnotationer, nemlig den romantiske drøm om hvid jul:

- (9) tunge skyer hang lavt og varslede sne. Måske ville det endelig blive en hvid jul? (Hjemmet, 1988)

Et andet eksempel på en atypisk kontekst mødte jeg ved forkortelsen for *kilometer*: *km*. Hvis man bliver forblændet af trangen til at oplyse, fristes man måske til at bringe følgende citat som belæg på forkortelsen *km*:

- (10) I stedet for at sige 1000 m siger vi 1 kilometer og skriver 1 km (Mogens Nørgaard Olesen: *Alt er tal*, 1990)

Som citat er dette helt uegnet, dels fordi det slet ikke giver nogen supplerende oplysninger (artiklen oplyser jo naturligvis at *km* er forkortelse for *kilometer*), dels fordi den lærebogsagtige formulering (som netop også er fra en skolebog) er helt atypisk sprogbrug i forbindelse med denne forkortelse. Meget mere naturlige og typiske citater vil være som de følgende:

- (11) Jeg løb ca. 7 km på 45 min, hvilket jeg tror er nær en rekord (dagbog, 1992)

- (12) den tilladte hastighed er 50 km i timen (Jyllands-Posten, 1992)

Fejl

Det er et meget kontroversielt spørgsmål om ordbogseksempler må indeholde fejl. Hvis der kun bringes redaktionelle eksempler, bør problemet ikke melde sig da redaktøren jo bare kan lade være at lave fejl! Anderledes forholder det sig med autentiske eksempler. Det var fristende at sige at når der nu er tale om autentisk sprog, så afspejler man bare det ubearbejdede sprog som det foreligger, og det er som sådan ofte fejlbehæftet, og redaktionen står ikke inde for det. Denne holdning er imidlertid ikke gangbar, specielt når man betænker den normerende virkning og de mange læsere som en ordbog som Den Danske Ordbog formentlig vil få. Det store flertal af brugerne vil uden tvivl forvente at finde korrekt sprogbrug i eksemplerne, sprogbrug som de vil kunne efterligne uden at få problemer med dansklærere, arbejdsgivere eller andre. Det giver os store problemer som ordbogsredaktører fordi en væsentlig del af alle eksempler på et givet ord indeholder større eller mindre sproglige eller faktuelle fejl.

Disse fejl kan så være mere eller mindre graverende. For at overholde vores målsætning om at være deskriptive (men altså

også for at få citategnede eksempler nok at vælge imellem!) tillader vi bl.a. følgende typer fejl i citaterne:

- kommafejl,
- forkert sær- og sammenskrivning (*idag* over for det normerede *i dag*),
- forkerte forkortelser (*kr* uden punktum; *TV*, som normeres *tv*),
- diskutabel brug af bindestreg (*gen-transplantation*; *latin-prøve*),
- forkert brug af apostroffer (*Kirsten's computer*).

Disse typer fejl anser vi altså for at være under bagatelgrænsen og er ikke bange for at lade dem slippe igennem i citater.

Hvis der ellers er stavfejl (både i og uden for opslagsordet), må vi gå uden om det pågældende citat. Der er således utallige tilfælde hvor – ellers gode – citater må falde fordi de indeholder en stavfejl. Jeg skal nøjes med at give et enkelt eksempel, nemlig til ordet *indgroet* i overført betydning:

- (13) Alt for ofte lader vi os styre mere af indgroede vanemønstre end af egentlige refleksioner over vort liv og forhold til andre mennesker (Massage Meditation Yoga, Center for bevidsthed og vækst, 1991)

Refleksion staves på dansk med -s-, så citatet, der ellers på glimrende vis modstiller vanemønstre og refleksion, må droppes, måske til fordel for et ringere citat – uden fejl!

Med hensyn til fejltyper der ligger uden for *Retskrivningsordbogens* område, altså den ortografiske norm, er situationen lidt vanskeligere. Selvom man som redaktør ønsker at være deskriptiv, går man i sit valg af citater mere eller mindre bevidst efter citater helt uden fejl eller kritisabel sprogbrug. Ej heller er man særlig tilfreds med at bringe et citat der indeholder en oplysning som man positivt ved er forkert (fx et forkert årstal). Eks. 14 indeholder en almindelig, men uheldig konstruktion (belæg på ordet *vandføring*):

- (14) Med en vandføring på ti kubikmeter pr. sekund stejler den lille gummi-båd i vandfaldet (B.T., 1991)

Verbalsubstantivet *vandføring* knyttes her til et forkert logisk subjekt, idet det selvfølgelig er vandfaldet der har en vandføring, og ikke gummibåden. Til trods for citatets åbenlyse kvaliteter (korthed, angivelse af vandmængde, sammenhæng med vandfald) vil vi føle os nødsaget til at droppe det for ikke at gengive noget der regnes for dårligt sprog.

Tilsvarende forholder det sig med talesprogscitater. Med alle de fænomener der er karakteristiske for spontan tale: tøvefænomener, selvrettelser, gentagelser m.m., er udskrevet talesprog meget svært at håndtere og bringe som citat uden at det kommer til at se meget mærkeligt ud for nogen der ikke er vant til at se den slags. Det følgende eksempel stammer fra en udskrevet samtale med en ung pige (belæg på ordet *rideskole*):

- (15) ude på rideskolen hvor jeg rider der er både {pause} altså {pause} privat {pause} altså {tøven} privatheste, ikke? og så rideskolens (Projekt Bysociolingvistik, 1987)

Der er her klart tale om spontant talesprog som vi alle sammen kender det, men det er kun sprogfolk der er klar over at det faktisk er sådan vi taler. Et eksempel af denne type kræver skrap redigering for bare at nærme sig de "normale" skriftsprogscitater, og selv da vil de fleste brugere nok stejle over det og ikke få det tilsigtede udbytte.

Tendentiøse eller anstødelige udsagn

Som omtalt kan vi som ordbogsredaktører til en vis grad fralægge os ansvaret for mindre sproglige bommerter i citaterne (kommafejl, bindestregsfejl e.l.). Men kan vi tillade os at viderebringe tendentiøse eller direkte anstødelige eller fornærmende udsagn under henvisning til at vi jo bare citerer autentisk sprog der opfylder vores formål, nemlig at illustrere brugen af et givet opslagsord? Lad mig give et par eksempler. Det første, som om-

handler offentlige skoler over for privatskoler, er vel egentlig ganske ufarligt selvom nogle nok vil føle sig stødt (belæg på *kommuneskole*):

- (16) da de københavnske kommuneskoler ikke har ry for at være særlig gode . . . har vi foretrukket at lade dig gå et sted hvor der er mere system i tingene [: en privatskole] (dagbog, 1986)

men langt værre bliver det når man kommer over i de racistiske og på anden måde groft diskriminerende udtalelser (belæg på ordene *jødesvin* og *perker*¹):

- (17) Nu har vi dig, dit kommunistiske jødesvin! (Bo Torstensen: *Sammen vi står!*, 1988)

og

- (18) Jeg hader at se perkerne skumme fløden imedens danskerne dør på grund af ventetid [i] hospitalssystemet (Indlæg til Fælleslisten mod indvandringen, Internettet 1997)

Ifølge Malmgrens artikel vil han tilsyneladende gerne være med til at optage enkelte af den slags opslagsord, men mener at det er etisk uforvarsligt at bringe racistisk betonedede eksempler som de ovenstående. Jeg mener at man principielt ikke bør afstå fra at illustrere et ords brug selvom man personligt tager afstand fra ordet og måske også fra de mennesker der bruger det. Det der er vigtigt, er selvfølgelig at give en kraftig advarsel om ordets stærkt nedsættende valør og netop kun bringe et sådant citat i den artikel der har det pågældende ord som opslagsord.

Manglende underholdningsværdi

Ud over at være eksempel på naturlig sprogbrug må et citat godt bringe billeder af konkrete situationer frem i læserens bevidsthed. Citaterne kan være individualiseringer af de mere formaliserede oplysninger der er bragt tidligere i artiklen i form af definition,

¹ Ordet *perker* er en stærkt nedsættende betegnelse for en mørk indvandrer.

eventuelle synonymer, antonymer, valensoplysninger m.m. Derfor vil det kunne diskvalificere et citat hvis det slet og ret er for kedeligt. Et enkelt eksempel hvor et tørt og faktisk citat er skiftet ud med et mere levende (belæg på ordet *tørretumbler*):

- (19) Der kom forslag om opsætning af møntautomat til betaling af det øgede strømforbrug i forbindelse med tørretumbler (Referat fra beboermøde, over for 1986)
- (20) jeg ordner lige lidt vasketøj. Det gør jeg næsten hver dag. Skønt med tørretumbler, når det øser sådan ned (dagbog, 1992)

Karakteristisk nok stammer det sidste citat fra en meget vigtig kilde til ord fra dagligdagen, nemlig en lang række dagbøger der blev skrevet på Nationalmuseets foranledning den 2. sept. 1992. Herfra får vi et stort udvalg af ord, konstruktioner og situationer som stammer fra almindelige menneskers almindelige hverdag.

Intet citat

I mange tilfælde er vi ude for at der selv blandt mange belæg slet ingen er der egner sig til at blive bragt som citat. Det gælder fx for mange ord med meget konkret betydning at belæggene ofte er temmelig intetsigende og kun kan tjene til dokumentation. Måske indgår ordet mest i opremsninger som ikke siger meget om ordet selv, eller måske står det som opslagsord i et leksikon der er blandt kilderne i korpus. I så tilfælde kan man lige så godt undvære et citat, medmindre formålet er for enhver pris at dokumentere at ordet findes i ens korpus.

En gruppe ord hvor det især er vanskeligt at finde citater, er nationalitetsord (*nordmand, norsk*). Ofte, især ved små og fjerne nationer, er ordene knap nok repræsenteret i vores korpus; i andre tilfælde er der masser af eksempler at tage af, men i så fald er de fleste belæg enten fulde af forældede fordomme, eller også stammer de ensidigt fra krigs- eller sportsreportager. Her er man ofte nødt til at lade være med at bringe noget citat selvom ordets

forekomsttal absolut berettiger til det, eller bringe et farveløst og politisk korrekt citat som fx:

- (21) Alt tyder på, at der ikke bliver rokket ved regeringens aftale med svenskerne om Øresundsforbindelsen (B.T., 1991)

En maliciøs sjæl havde måske lyst til at drille broderfolket med citater af typen:

- (22) Fuld svensk vægtløfter mørbankede to politifolk (Berlingske Tidende, 1991)

men der er jo ingen grund til at sætte det gode naboskab over styr!

Afslutning

Jeg har i det foregående givet eksempler på hvor vanskeligt det er at finde citategnede belæg i vores store tekstkorpus. Det ideelle citat skal, som det er fremgået, imødekomme en række krav som jeg kort skal resumere

- det skal være tilpas kort;
- det må ikke indeholde forstyrrende elementer der tager opmærksomheden bort fra opslagsordet;
- det må ikke indeholde for svære ord;
- det skal være naturligt med hensyn til sprog og situation;
- det må ikke indeholde fejl;
- det bør kun indeholde anstødelige udsagn hvis man gør udtrykkeligt opmærksom på stilværdien;
- det må gerne være underholdende og levende.

Disse krav er svære at opfylde på en gang, og man må i mange tilfælde stille sig tilfreds med et mindre godt citat eller acceptere forstyrrende afkortninger, indskudte forklaringer e.l.

Hvorfor kan I dog ikke bare rette fejlene og skære citaterne til så de passer, uden at gøre rede for det? vil nogen måske spørge. Dertil er svaret: fordi vi lægger os i ODS-traditionen med hensyn til videnskabelighed. Vi citerer korrekt, og hvis vi ændrer noget i et citat, skal det fremgå. Desuden skal citaterne være forsynet med kildehenvisning så brugeren kan se hvor og hvornår den pågældende sætning blev skrevet eller sagt. Til gengæld må vi så leve med at der bliver lidt ”støj” i form af prikker og parenteser, men disse gener bliver i højeste grad opvejet af fordelene ved de autentiske eksempler, nemlig det nye og spændende som de kan tilføre ordbogen, modsat de redaktionelle eksempler.

Litteratur

- Bergenholtz, Henning. 1994. ”Beispiele in Fachwörterbüchern.”
Burkhard Schaedler/Henning Bergenholtz (Hrsg.): *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern*. Tübingen: Gunter Narr Verlag. S. 321–439.
- Malmgren, Sven-Göran. 1994. ”Språkprovens form og funktion i svenska betydelseordböcker från Östergrens Nusvensk ordbok till Svensk ordbok.” *LexicoNordica* 1:107–117.