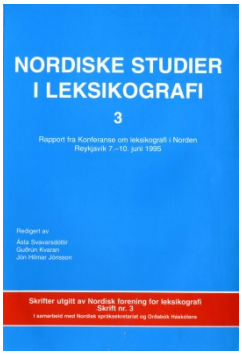
NORDISKE STUDIER I LEKSIKOGRAFI

Titel: Om udvælgelse af ord til Den Danske Ordbog

Forfatter: Henrik Andersson

Kilde: Nordiske Studier i Leksikografi 3, 1995, s. 1-9
Rapport fra Konferanse om leksikografi i Norden, Reykjavík 7.-10. juni 1995

URL: http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive



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 Andersson

Om udvælgelse af ord til Den Danske Ordbog

The Danish Dictionary is to contain approximately 100 000 entries selected from a text corpus of 40 million running words and four comprehensive existing dictionaries. A method of automatic selection was developed on the basis of two entry lists, which were "manually" selected from small intervals of the alphabet. Two similar entry lists were then selected computationally. The linguistic importance of a word was calculated on the basis of the number of corpus instances and the number of existing dictionaries in which the word occurs. This led to the formulation of some 30 criteria for the automatic selection of entries.

1 Indledning

Det følgende er en praktisk-leksikografisk beskrivelse af den arbejdsproces, der har ført frem til den lemmaliste, som 10 redaktører ansat på *Den Danske Ordbog* (herefter DDO) bruger i det daglige redigeringsarbejde.

Inden man giver sig i kast med at vælge ord ud, bør tre spørgsmål være besvaret:

1. Hvilken type ordbog er der tale om?
2. Hvor mange og hvilke oplysninger skal ordbogen give?
3. Hvilke kilder har redaktørerne til rådighed?

Historien har vist, at det ikke er nogen selvfølge, at disse vigtige spørgsmål er afklaret, inden ordudvælgelse og artikelskrivning går i gang. For DDO's vedkommende lyder svarene:

2 Ordbogens art

DDO er en almensproglig, monolingval betydningsordbog, der skal beskrive det danske sprog i perioden fra ca. 1950 til i dag, med hovedvægten på tiåret 1983–92 (primærperioden). Ordbogen skal dække det skrevne sprog og inddrage det talte og henvender sig i første række til brugere med dansk som modersmål, i anden instans til personer med dansk som andet- eller fremmedsprog. Den skal indeholde ca. 100 000 opslagsord. Ordbogen kommer til at fylde seks bind a ca. 100 000 linjer.

3 Oplysningstyper

Artiklerne skal give oplysning om:

- ortografi
- udtale
- bøjning
- betydning
- kombinatorik (fx kollokationer og valensmønstre)
- orddannelse (afledning og sammensætning)
- etymologi

4 Ordbogens kilder

1. Et elektronisk tekstkorpus på 40 mio. løbende ord fra primærperioden 1983–92.
2. Eksisterende ordbøger og leksikaliserede ordsamlinger, alle i nyeste udgave i maskinlæsbar form, nemlig:
   - *Retskrivningsordbogen*, udgivet af Dansk Sprognævn (RO; ca. 60 000 opslagsord).
   - Blinkenberg & Høybye: *Dansk-fransk ordbog* (B&H; over 150 000 opslagsord).
   - Vinterberg & Bodelsen: *Dansk-engelsk ordbog* (V&B; over 150 000 opslagsord).
   - Dansk Sprognævns register (DS) over især sammensætninger og afledninger i dansk fra 1950 og frem.

5 Grunde til at udvælge hele ordstoffet inden redigering

DDO tager over, hvor *Ordbog over det Danske Sprog* (ODS; i 28 bind + 5 supplementsbind, der er under udgivelse) kronologisk hører op. Men DDO er på mange måder en helt anden ordbog. For det første har den meget mindre plads til rådighed, for det andet dækker ordbogen en meget kortere periode (ca. 50 år mod ODS' ca. 250 år). Og for det tredje er DDO's arbejdsredskaber nogle helt andre end ODS'.

Det stod klart fra første færd, at redaktionen ikke kunne medtage „det hele", som ODS + Supplement. Med et koncept, der kun stiller 600 000 ordbogslinjer til rådighed, måtte ordene udvælges med skønsomhed. Det blev hurtigt besluttet, at hele ordstoffet om muligt skulle ligge klar, inden redigeringsarbejdet for alvor gik i gang.

Herved ville der blive mulighed for at redigere i semantisk relaterede ordfelter, fx 'nedsættende betegnelser for mandspersoner': *bisse*, *bølle*, *kanalje*, *laban*, *quisling*, *sjuft*,

skurk osv. Det er nemmere og hurtigere at redigere den slags ordfelter på én gang end at behandle dem, efterhånden som man møder dem på sin vej gennem alfabetet. Dertil kommer, at det håndværksmæssigt er mere forsvarligt: Det bliver nemmere at opfange og beskrive hårfine betydningsnuancer mellem næsten-synonymer og chancerne for konsistent angivelse af synonymi, antonymi o.l. øges.

En anden fordel, man opnår ved at have hele ordstoffet til rådighed, er, at man kan foreskrive sig, ja måske ligefrem overholde den regel, at simpleksord, der bruges i betydningsdefinitioner, skal kunne slås op i ordbogen.

Sidst, men ikke mindst følger der en administrativ gevinst med en tidlig ordudvælgelse. Man kan løbende kontrollere, om artiklerne overholder det omfang, der er fastlagt i planen.

6 Udvikling af en metode til automatisk udvælgelse

Man kunne umiddelbart forestille sig to yderligtgående måder at vælge ord ud på. Som det ene ekstrem kunne man samle alle forskellige ord fra korpus og eksisterende ordbøger i en liste og gennemgå dem et for et. Det bedste, man kan sige om den metode, er, at den er grundig. Men med omtrent 800 000 lemmakandidater til rådighed siger det sig selv, at den ville være alt for langsom. Man kunne også gå i den modsatte grøft og udelukkende bruge korpus til at udvælge ord fra, fx ud fra reglen: „forekommer et ord x gange i korpus, skal det med", hvor x = den værdi, der giver ca. 100 000 forskellige opslagsord. Fordelen ved denne metode ville være, at den var særdeles hurtig. Men også meget overfladisk og på anden måde uhensigtsmæssig. For det første ville der ikke blive taget højde for, at der er en indbygget diakron skævhed i korpusmaterialet, der jo kun dækker tiåret 1983–92, mens planen foreskriver, at ordbogen skal omfatte perioden helt fra 1950 til i dag. Ord fra den tidlige del af perioden ville blive mangelfuldt repræsenteret. For det andet kunne der være brist i forekomsten af fagsprog, også det mere almene. Der kan siges meget godt om den engelske *Collins Cobuild*-ordbog, hvis lemmaselektion er rent korpusbaseret. Men hvor ofte har man ikke forgæves forsøgt at slå almene fagord op i den! For det tredje har det vist sig, at korpus indeholder mange forholdsvis frekvente, men semantisk og på anden måde intetsigende ad hoc-sammensætninger og -afledninger, ord, som ikke har interesse i en betydningsordbog.

En metode til automatisk ordudvælgelse måtte manøvrere mellem Skylla og Karybdis. En lille procentdel af ordstoffet blev udvalgt efter den grundige, men langsommelige alfabetisk fremadskridende ord for ord-metode. Derefter skulle denne del af ordstoffet danne grundlag for fastlæggelsen af en automatisk selektion. Processen forløb i fire etaper:

1. Ord for ord-udvælgelse af 2% af ordstoffet (det alfabetiske interval *gal–greb*).

2. Ord for ord-udvælgelse af yderligere 4–5% af ordstoffet (hele bogstav *a*).

3. Computersimulering af de to samme alfabetiske udsnit med efterfølgende sammenligning.

4. Automatisk udvælgelse af resten af ordstoffet.

6.1 1. etape: *gal–greb*

Første etape blev afviklet i en tidlig fase af forløbet, hvor redaktionenen havde brug for hurtigt at få udvalgt en del af ordstoffet til omgående redigering. Det måtte afklares, om ordbogens planlagte omfang overhovedet kunne overholdes, når det kom til stykket, eller om antallet af oplysninger om ordene måtte skæres ned. At netop udsnittet *gal–greb* blev valgt som prøveklud, skyldes, at forholdet mellem små og store artikler i dette toprocentsinterval svarer ret nøje til de resterende 98% af alfabetet.

I denne tidlige fase af forløbet var kilderne endnu ikke gjort klar til automatisk udnyttelse. Metoden eller manglen på samme gik ud på, at to redaktører i fællesskab slog alfabetiske delintervaller op i korpus og talte korpusforekomster sammen, hvorefter de valgte ordene ud efter frekvens. RO og redaktørernes sprogfornemmelse supplerede med ord, som enten ikke fandtes, eller som var svagt repræsenteret i korpus.

Resultatet forekom ganske tilfredsstillende, men det havde også taget to fuldtidsansatte redaktører ca. en måned at udvælge de sølle 2% af ordstoffet. Det stod klart, at der måtte langt mere fart på lemmaselektionen; ellers ville der ved deadline i 1999 højst foreligge en ordliste, men ikke nogen ordbog.

Mens *gal–greb*-ordene blev udvalgt og redigeret, fik redaktionens to edb-eksperter, Jørg Asmussen og Ole Norling-Christensen, udviklet to vigtige redskaber til brug for arbejdet med udvælgelsen af hele bogstav *a*.

For det første blev samtlige ord i korpus forsynet med frekvensoplysninger, så man slap for det møjsommelige arbejde med at tælle korpusforekomster sammen i hovedet. Frekvensen blev angivet som to værdier: én for samlet antal forekomster og én for antal korpusdokumenter. Det er den sidste værdi, der tæller. I det følgende betyder „x korpusforekomster" altså 'forekommer i x forskellige tekster i korpus'. Optræder et ord fx 16 gange i den samme tekst, regnes det for én korpusforekomst.

For det andet blev samtlige ord fra ordbøgerne indlæst i en datafil, med angivelse af, hvilke(n) kilde(r) de stammede fra.

6.2 2. etape: bogstav *a*

I denne etape blev der lagt større vægt på ordbogskilderne end i *gal–greb*-udsnittet. Forekom et ord i alle ordbøgerne, altså både i RO, B&H, V&B og i Sprognævnets register, blev det udvalgt uden hensyn til, om det var repræsenteret i korpus. Ordene *adjunktur*, *agnosticisme*, *appellativ* og *attributiv* er de vistnok mest prominente eksempler på ord, der forekom i alle ordbøger, men ikke i korpus. Man kunne også nævne *akribi*, hvad der forhåbentlig ikke skal lægges noget symbolsk i!

Forekom et ord i de tre rigtige ordbøger — altså RO, B&H og V&B, men ikke i Sprognævnets register, stod det også på forhånd stærkt. Under udvælgelsen blev der skelet til korpusfrekvensen, men tommelfingerreglen lød: Er du i tvivl, skal det med.

Fandtes et ord kun i tosprogsordbøgerne B&H og V&B, stod det svagt. Tosprogsordbøger indeholder jo tit vendte fremmedsprogsækvivalenter, gamle ord, forældet fagsprog o.l., som kan have relevans i netop dén type ordbøger; men i DDO hører de ikke hjemme.

En del af ordene i Sprognævnets register var ikke overraskende ad hoc-sammensætninger og -afledninger, der aldrig har slået rod i det danske sprog. Ord, der kun var repræsenteret

i denne ordbogskilde, stod som hovedregel svagt, medmindre solid korpusfrekvens talte for det modsatte.

Efter at ordene fra de forskellige ordbogskombinationer var gennemgået og udvalgt, blev lemmalisten suppleret med højfrekvente ord fra korpus, som ikke var med i de eksisterende kilder. Hermed var der valgt lidt over 4000 bogstav *a*-ord ud, hvilket passede godt med, at dette afsnit af alfabetet erfaringsmæssigt fylder 4–5% af en ordbog.

Det samlede indtryk af de udvalgte ord var, at der var mange ord i en grå zone, ord, som måske/måske ikke skulle med. Navnlig forekom mange ord med lav korpusfrekvens problematiske. Frygten for, at korpus skulle mangle ord fra den tidlige del af perioden (før 1983) og almene fagord, havde givet de eksisterende kilder et lidt for stort ord at skulle have sagt. Ikke mindre end 20–25% af de valgte ord befandt sig i den grå zone. Her blot syv eksempler:

Heterograf	*Klasse*	*Korp. forek.*	*Lexkilder*
afsidning	sb	0	KUN B&H+V&B
afsikring	sb	0	KUN RO
afskedigelsesløn	sb	0	ALLE
afskedigelsesnævn	sb	0	IKKE RO
afskibningshavn	sb	0	RO,B&H,V&B
afskrue	vb	0	KUN B&H+V&B
afskrå	vb	0	KUN B&H+V&B

Ordet *afskedigelsesløn* blev automatisk udvalgt ud fra reglen om, at ord, som fandtes i alle kilder, skulle med, også selv om de ikke var repræsenteret i korpus. Eksemplet viser, at reglen ikke er skudsikker. Inden for primærperioden bruges *afskedigelsesløn* stort set ikke, det hedder i dag med en tidstypisk eufemisme *fratrædelsesgodtgørelse* (8 forekomster i korpus); men *fratrædelsesgodtgørelse* er så svagt repræsenteret i ordbøgerne, at det ikke ville komme med efter de omtalte selektionsprincipper. Eksemplet viser, at overdreven tillid til de eksisterende ordbøger kan medføre en overtrædelse af planens forskrift om, at hovedvægten i DDO's sprogbeskrivelse skal ligge på perioden 1983–92.

Ordene *afskedigelsesnævn* og *afskibningshavn* er også velrepræsenterede i ordbøgerne. Ret beset er det imidlertid diskutabelt, om de hører hjemme i DDO. Sammensætningerne *afskedigelsesnævn* og *afskibningshavn* er semantisk transparente (*afskibe* og *afskibning* er velbelagt både i korpus og i ordbøgerne), og ingen af ordene har vel været særlig udbredte på noget tidspunkt i perioden, heller ikke den tidlige (ODS-Supplementet har ét belæg på *afskibningshavn*, fra 1948, men ingen på *afskedigelsesnævn*). Korpus viser, at *-nævn* som andet sammensætningsled er produktivt i primærperioden, så det er ikke indlysende at vælge *afskedigelsesnævn* med 0 forekomster, bare fordi RO, V&B og B&H har ordet med. Sammensætninger som *adoptionsnævn*, *aftalenævn*, *forældrenævn*, *lønningsnævn* og *ungdomsnævn* er alle velrepræsenteret i korpus, men står så svagt i de eksisterende ordbøger, at de ikke ville få et ben til jorden efter de her skitserede udvælgelseskriterier.

At verbalsubstantiverne *afsidning* og *afsikring* kom med trods svag kilderepræsentation, skyldes, hvis sandheden skal frem, snarere en subjektiv fornemmelse end saglige forhold.

At partikelverberne *afskrue* og *afskrå* blev valgt ud trods svag ordbogsrepræsentation, er endnu dårligere begrundet. V&B og B&H følger begge princippet om, så vidt muligt at lade den sammensatte form være indgang til partikelverber. Redaktøren, der stod for

udvælgelsen, har elimineret så mange partikelverber med *af-* fra de to tosprogsordbøger, at han er blevet nervøs for, om der nu kom nok bogstav *a*-ord med. Derfor har han medtaget nogle af dem, der umiddelbart har forekommet ham mindst urimelige.

Anden fase viste, at man ikke burde lægge for megen vægt på de eksisterende ordbøger. For mange afgørelser kom til at bero på subjektive skøn. Spørgsmålet var nu, om et edb-program kunne udnytte erfaringerne og præstere et mere plausibelt resultat.

6.3 3. etape: edb-simulering af *gal–greb* og bogstav *a*

Tredje etape gik groft sagt ud på at besvare spørgsmålet: Hvordan får man et edb-program til at ramme de menneskeudvalgte ord så præcist som muligt?

Fremgangsmåden, som blev udviklet af redaktør Jørg Asmussen, var som følger:

Først blev en lille del af det oprindelige *gal–greb*-afsnit skilt ud, nemlig intervallet *gallicisme* til *gammel* (48 ord, ca. 2 promille af det samlede ordstof). For at få alle 48 ord valgt ud automatisk, viste det sig, at tre kriterier skulle være opfyldt, nemlig:

1. 4 *korpusforekomster + repræsentation i mindst 3 af de 4 ordbøger.* De resterende kandidater skulle

2. *mindst være repræsenteret i 3 ordbøger, uanset korpusfrekvens.* De sidste ord blev opfanget ved opstilling af kriteriet

3. *mindst 5 korpusforekomster uanset repræsentation i ordbøgerne.*

De tre selektionskriterier blev derefter overført på hele bogstav *a*. Ikke uventet viste det sig, at kriterierne var for grove; der kom alt for mange irrelevante ord med. Derfor blev der opstillet nogle finindstillingskriterier, der så vidt muligt skulle beholde de relevante, allerede udvalgte bogstav *a*-ord, men udskille de irrelevante. Det ville føre for vidt, her at anføre samtlige finindstillinger, men som eksempler kan nævnes, at ord med mindre end fire korpusforekomster og repræsentation i kun én af tosprogsordbøgerne (B&H el. V&B) blev elimineret. Ligeledes ord med mindre end to korpusforekomster og repræsentation i B&H og V&B, men ikke andre ordbøger.

Herefter blev de mere fintmærkende selektionskriterier overført på et større afsnit af *gal–greb*-ordene. En række nye kriterier blev defineret, atter overført på hele bogstav *a* og så fremdeles, indtil der var et maskinelt udvalg af *gal–greb-* og bogstav *a*-ord, der kom så tæt som muligt på ord for ordudvalgene af de samme alfabetiske intervaller.

Det viste sig umuligt at lave en helt homogen edb-simulering. Nogle få korpus/ordbogskombinationer resulterede i ordlister, der på én gang indeholdt oplagt urimelige og yderst relevante lemmakandidater. Derfor blev ordene inddelt i tre kategorier:

Kategori a): Ord fra både korpus og eksisterende ordbøger, der tilhører det centrale ordforråd og skal have selvstændig indgang, ca. 75% af ordstoffet.

Kategori b): Ord fra kombinationer med uhomogene lemmakandidater. Det er op til den enkelte redaktørs skøn, om det enkelte kategori b-ord skal med som selvstændig indgang, skal degraderes til eksempel på sammensætning el. afledning eller evt. skal smides helt ud. Til denne kategori hører bl.a. alle ord med stort begyndelsesbogstav,

der jo ofte, men ikke altid er proprier, fx *A-dur, AIDS, ATP-bidrag*. Kategori b-ordene kom til at udgøre ca. 20% af det samlede ordstof.

Kategori c): Ord, der ikke findes i ordbogskilderne, men har en rimelig korpusfrekvens (fem forekomster el. mere). Kategorien repræsenterer på én gang nogle af de interessanteste ord (neologismer), og nogle af de mest uinteressante (banale ord som fx sammensætninger med *-forretning*, *-sag* og *-situation* som andetled, afledninger med *-mæssig*, blot for at nævne et par stykker).

Den samlede gennemgang af ord fra de to udvælgelsesmetoder gav et plus til den automatiserede. Gråzoneordene blev elimineret, og opdelingen af ordstoffet i de tre omtalte kategorier overlod trods alt et vist initiativ til den menneskelige dømmekraft. Som forventet var ingen af metoderne helt skudsikre. De efter bedste skøn vigtigste bogstav *a*-ord, som den automatiserede udvælgelse ikke fik med, var:

Heterograf	*Klasse*	*Korp. forek.*	*Lexkilder*
adjunktur	sb	0	ALLE
afhentningspris	sb	0	RO + DS
afrofrisure	sb	0	RO + DS
afspændingsmiddel	sb	1	RO + DS
agility	sb	1	INGEN
agitprop	sb	2	KUN DS
amnesi	sb	0	RO,B&H,V&B
andengenerationsindvandrer	sb	1	INGEN
antiroman	sb	0	V&B,B&H,DS
armyjakke	sb	3	KUN DS
artistnummer	sb	5	INGEN

Af disse ord er *agitprop* og *antiroman* eksempler på ord, der havde større udbredelse i sproget i den del af perioden, som ikke er dækket af korpus. En overraskende stor del af disse ord blev reddet vha. den betingelse, at ord, der forekom i RO, B&H og V&B blot skulle forekomme én gang i korpus for at blive klassificeret som kategori a-ord.

Langt flere af ordene har først fået borgerret i sproget efter primærperiodens udløb: *afhentningspris*, *afrofrisure*, *afspændingsmiddel*, *agility*, *andengenerationsindvandrer* og *armyjakke*.

At et ord som *adjunktur* savnes, kan skyldes, at ordbogsredaktører typisk er fortrolige med undervisnings- og universitetsmiljøet. Hvis der er tale om en usaglig præference, ligger den helt på linje med, at sprogvidenskabelige termer ofte prioriteres højere i almensproglige ordbøger end så mange andre nok så relevante fagområder. Det er vist først og fremmest sprogvidenskabsmænd og ordbogsredaktører, ikke den almindelige bruger, der vil savne tidligere omtalte ord som *appellativ* og *attributiv*.

Værre er det nok med et så relativt almindeligt medicinsk fagudtryk som *amnesi*, der mangler den ene sølle forekomst, der kunne have reddet ordet.

Listen antyder et problem, som det ville føre for vidt at komme ind på her, nemlig edb-støj. Ordet *artistnummer* findes ikke i de eksisterende ordbøger, hvilket medfører, at programmet ikke har kunnet lemmatisere ordet. Bøjningsformen *artistnumre* tegner sig for

fire af de fem forekomster, men er af programmet opfattet som et andet ord end *artistnummer*, der har én forekomst. Ordet er dermed blevet elimineret ud fra reglen om, at et ord, der ikke forekommer i de eksisterende ordbøger, skal forekomme mindst fem gange i korpus for at blive udpeget.

Den langsommelige ord for ord-udvælgelse af bogstav *a* lagde større vægt på kilderne; men også den havde sine smuttere, hvoraf de 11 grelleste er:

Heterograf	*Klasse*	*Korp.forek.*	*Lexkilder*
ABS-bremse	sb	47	KUN DS
adresseliste	sb	8	KUN V&B
advokatsalær	sb	7	KUN V&B
afbudsrejse	sb	7	IKKE B&H
aha-oplevelse	sb	6	KUN V&B
almenvel	sb	30	RO + B&H
altmodisch	adj	9	IKKE V&B
ankelsok	sb	18	RO + B&H
appelsinsaft	sb	60	KUN B&H
artisteri	sb	16	RO + V&B
ayatollah	sb	34	RO + B&H

Som det fremgår, er der tale om ord med pæne frekvenstal (ml. 6 og 60 forekomster i korpus). *ABS-bremse* er så nyt et ord, at det af naturlige grunde ikke er kommet med i de udgaver af ordbøgerne, som fandtes på udvælgelsestidspunktet. Derimod forekommer det overraskende, at V&B er ene om at have *adresseliste*, *advokatsalær* og *aha-oplevelse* med, og at kun B&H har fundet *appelsinsaft*, listens topscorer mht. korpusfrekvens, værdigt til optagelse. I de øvrige tilfælde: *afbudsrejse*, *almenvel*, *altmodisch*, *ankelsok*, *artisteri* og *ayatollah* er det en enkelt af de to tosprogsordbøger, der har fravalgt ordet. Det skal her understreges, at de valgte eksempler ikke er anført for at kritisere hæderkronede ordbøgers lemmaselektion. Tværtimod må man anerkende, at der er så få tilfælde af uenighed om, hvad det centrale ordforråd er.

7 Konklusion

Der er betydelige fordele ved at automatisere udvælgelsen af ord. Det er ikke alene den hurtigste (og dermed billigste) løsning, men også den bedste, for så vidt som den giver det mest konsistente udvalg. Naturligvis er sprogfornemmelsen en faktor, man ikke kan se bort fra, men når det kommer til konkrete afgørelser af, hvilke ord der skal med i en ordbog, og hvilke der ikke skal, er det næsten uundgåeligt, at en vis vilkårlighed — subjektive præferencer og idiosynkrasier — gør sig gældende. Den beskrevne udvælgelsesmetode giver en saglig begrundelse for optagelsen eller forkastelsen af hver eneste af de omtrent 800 000 lemmakandidater, kilderne tilsammen indeholder.

Litteratur

B&H = Andreas Blinkenberg/Poul Høybye 1991: *Dansk-fransk ordbog*. 4. udg. ved Jens Rasmussen & al. Bind 1–2. København: Nyt Nordisk Forlag Arnold Busck.

DS = Dansk Sprognævns register over sammensætninger og afledninger i dansk 1950 og frem. Ikke publiceret.

Kristensen, Kjeld 1993: Den Danske Ordbogs tekstkorpus og spORDhunde. I: Anna Garde/Pia Jarvad (udg.): *Nordiske Studier i Leksikografi* II. København: LEDA/Nordisk Forening for Leksikografi, 138–42.

ODS = *Ordbog over det Danske Sprog*. 1918–56. Bind 1–28. København: Det Danske Sprog- og Litteraturselskab.

RO = *Retskrivningsordbogen*. 1986. København: Dansk Sprognævn.

V&B = Hermann Vinterberg/C. A. Bodelsen 1990: *Dansk-engelsk Ordbog*. 3. udg. ved Viggo Hjørnager Pedersen. København: Gyldendal.