

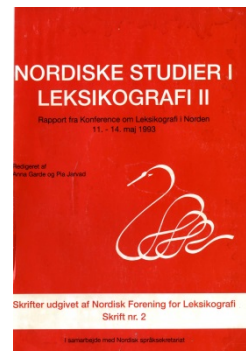
NORDISKE STUDIER I LEKSIKOGRAFI

Titel: Om datering af ordene i dansk

Forfatter: Anne Duekilde

Kilde: Nordiske Studier i Leksikografi 2, 1993, s. 53-68
Rapport fra Konference om leksikografi i Norden, 11.-14. maj 1993

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.

Om datering af ordene i dansk

Anne Duekilde

Det hører med til almindelig nysgerrighed og interesse for sproget at spørge: hvor længe har vi sagt sådan? det føles ikke ligegyldigt om en ordanvendelse er 2 år eller 20 år eller 200 år gammel. For forskningen har oplysninger om ordenes levetid i sproget både sproghistorisk og kulturhistorisk interesse; det kan være nyttigt at vide hvilket sprogligt materiale der var til rådighed på et bestemt tidspunkt, og oplysningerne om en bestemt epokes nye ord siger noget om kulturindflydelse og om epokens sproglige aktivitet.

Ved visse ordbogstyper er datering det centrale emne. Det gælder ordbøger over nye ord i en vis periode og de sjældnere ordbøger over ord der forældes i en vis periode. Dateringsforskningen kan resultere i en kronologisk ordbog, som viser forandringerne i ordforrådet gennem en vis periode¹. Dateringsspørgsmålet indgår også i de store nationalordbøger, og ud fra deres materiale har forskellige mindre ordbøger bragt dateringsoplysninger.

Da min baggrund for at vove mig ind på dette omfattende emne er arbejdet som redaktør ved Supplement til Ordbog over det danske Sprog (ODS-S.Iff. 1992ff.), vil jeg specielt se på hvordan (begyndelsen af) ODS og ODS-S bidrager til dateringen af ordene; desuden vil jeg berøre forholdet mellem etymologisk og semantisk datering samt visse problemer ved førstebelæg og sidstebelæg.

ODS som dateringskilde

ODS er planlagt som en stor historisk citatordbog der skulle blive hurtigt færdig, nemlig på cirka 20 år. ODS ville i 15-17 bind beskrive samtidssproget og desuden vise hvordan dette havde udviklet sig siden 1700. Ambitionsniveauet blev lagt betydeligt lavere end i Svenska Akademiens Ordbok, som var nået frem til bogstav D i de år begyndelsen af ODS blev redigeret. Første halvbind af ODS udkom 1918, og halvdelen af alfabetet blev faktisk redigeret på cirka 10 år. Sidste egentlige ordbogsbind udkom i 1954 og det afsluttende forkortelsesbind i 1956; det blev til i alt 28 bind. Den ældre nydanske tid, 1500-1700, behandles ikke i ODS; denne periode havde netop fået en leksikografisk behandling i Otto Kalkar, Ordbog til det ældre danske sprog (1300-1700), hvis store supplementsbind, værkets bind V, blev afsluttet i 1918.

ODS er både en samtidsordbog og en ordbog over de to foregående århundreder. Fra hele perioden blev der gennemgået tekster af mange slags, men interessen var ikke lige-

ligt fordelt mellem fortid og nutid. Nutidssproget var hovedsagen, og derfor blev der oprindelig ikke lagt særlig vægt på ordenes datering. Det var dog en fast regel at det ældste kendte belæg på et ord eller en ordanvendelse altid blev medtaget. De skønlitterære hovedforfattere lige fra Ludvig Holberg til Andersen Nexø blev godt repræsenteret i seddelsamlingen, men fra den ældre tid blev forholdsvis få andre tekster gennemgået, hvorimod samtidens bedste opslagsværker inden for alle fagområder blev udnyttet. I den senere del af ODS blev der inddraget flere ældre tekster.

Dateringsoplysninger kan i ODS fremtræde enten som en redaktionel oplysning eller i form af kildemateriale. Man kan skelne mellem 3 slags redaktionelle oplysninger:

1. Oplysninger om forekomst i en sprogperiode før 1700.
2. En redaktionel dateringsoplysning ved ord eller ordanvendelser der først er kendt i ODS' periode, fx: *I. Ballade*, digt: "laant i slutn. af 18. aarh. fra ty. ballade", *II. Ballade*, løjer: "optaget i kbh. i slutn. af 1860'erne fra parisisk ballade, spadseretur uden bestemt maal, kun for at drive ell. more sig", *Barndoms-*: "først alm. i 19. aarh., jf. at VSO.(1793)² ingen eksempler har", *benytte*: "dannet i 1. halvdel af 18. aarh. efter ty. benutzen, benützen, af ældre nytte".
3. Et vinkeltegn "<" foran forkortelsen for en ordbog er en indirekte dateringsoplysning. Det angiver at ordet (eller ordanvendelsen) ikke er medtaget i vedkommende ordbog; det er endda forklaret med ordene: yngre end, i tillid til at ordbøger altid afspejler den eksisterende sprogbrug. Fx: *afklare* 2, bringe klarhed m.h.t. noget: < VSO.1793, *Alvorsmand*: < MO.1859 (førstebelæg 1881, ODS-S 1818), *aflaste*: < S&B.1883, *berolige*: < vAph.(1764).

Hovedreglen for det nye ordstof fra ODS' periode 1700ff. er at dateringen må aflæses af kildematerialet. Det samme gælder for det ordstof som er gået af brug i løbet af perioden; yngste belæg må findes i kildematerialet. Det er derfor en alvorlig mangel ved ODS at der ikke er sat årstal på ethvert belæg; og denne mangel kan et supplement ikke afhjælpe. Læseren får altså normalt ikke mulighed for at aflæse ordets alder eller levetid direkte, men skal i reglen gå omvejen over forkortelsesbindet. Hvis citatet er fra en udgave hvis trykår er meget forskelligt fra tekstens år, fx samlede eller udvalgte værker, kan læseren endda stå uden anden datering end forfatterens fødsels- og dødsår.

ODS er en betydningsordbog. Kildematerialet er derfor grupperet betydningsmæssigt, og under hver gruppe er det ordnet kronologisk. Man skal se sig godt for, hvis man vil finde det ældste belæg på ordet som helhed, for det står ikke altid tidligt i artiklen, og det kan stå i en underordnet citatgruppe.

Trods de nævnte forhold er ODS vor vigtigste kilde for dateringsoplysninger inden for den periode ordbogen skal dække.

ODS-S som dateringskilde

ODS-S har til formål at supplere ODS' oplysninger om det danske sprog i perioden 1700-1955. Modsat ODS, og de store udenlandske ordbøger som var forbilleder for

ODS, har ODS-S altså sproget i en afsluttet periode som sin genstand, og ikke det dagsaktuelle sprog. Holdningen til det 20. århundrede bliver derfor tilbageskuende, og et "nu" indgår ikke i redaktionssproget. Så vidt det er muligt bliver de supplerende oplysninger fremstillet i de redaktionelle rammer som er givet i ODS. Dette gælder både for de artikler der supplerer en ODS-artikel og for de helt nye artikler. I ODS-S I er der 10.356 opslagsord, og cirka halvdelen supplerer en ODS-artikel. I lidt over 500 af disse står der et "+ ældre: "; på denne måde er et sprog- eller kulturhistorisk markant ældre belæg gjort synligt.

En gennemgang af disse tilbagedateringer inden for ODS' periode viser en klar dominans af tilfælde hvor ODS har ældste belæg omkring 1900. Her i begyndelsen af alfabetet, ODS-S I går fra A (inklusive Aa) til Bh, er ODS i allerhøjeste grad samtidsordbogen - som ganske vist var historisk, men som ikke var i besiddelse af tilstrækkeligt med informationer om fortiden. Ældste belæg er påfaldende ofte Salmonsens Konversationsleksikon, 1. udg. 1893-1911; eller det er fra tidens bedste faglitteratur inden for landbrug, teknik, handel, medicin osv. Selv i tilfælde hvor ældre opslagsværker eller ældre udgaver af benyttede værker kunne have været efterset, har man oftest ladet sig nøje med samtidens standardværker. Til en vis grad findes denne svaghed gennem hovedparten af alfabetet, specielt er fremmedordene i alt for høj grad forbigået ved excerperingen til ODS.

Tilbagedateringer fra ca. 1900 til det anførte år: *Aarebrok* (> 1810), *Aarring* (i en træstamme, > 1858), *Afgangsprøve* (> 1856), *Akvarium* (> 1858), *Allitteration* (belæg fra sprogforskeren Otto Jespersen, > 1816), *ambulatorisk* (> 1850), *Anlægs kapital* (> 1851), *anomal* (> 1838), *Association* (filos., psyk. > 1824), *Bakterie* (> 1868), *Barometerstand* (> 1826), *Begonia* (> 1834), *Benkul* (> 1861), *Beskyttelsestold* (> 1840), *Bevægenerve* (> 1840). Nogle eksempler fra det mere almene tale- og skriftsprog: *Abort* (billedligt om et mislykket åndsprodukt, > 1845), *akademisk* (i betydningen formel, regelbunden, > 1735), *Akkordarbejder* (> 1816), *Akrobat* (> 1828; ODS oplyser, vel for at bøde på sit utilfredsstillende kildemateriale: "først brugt i 19. aarh."), *altomfattende* (> 1811), *amerikanisere* (> 1833: et amerikaniseret Europa), *analytisk* (> 1775), *Anneksion* (om en stats indlemmelse af et landområde, > 1864) *annekttere* (om en stat, > 1870), *annoncere 2* (bekendtgøre gennem annoncer, > 1825), *appetitvækkende* (> Søren Kierkegaard, 1840'erne), *Arbejderurolighed* (> 1848), *arbejdsbesparende* (> 1857; jf. *arbejdssparende* 1859 > 1827), *Arrigtrold* (> 1810), *Badulje* (støj, spektakel, > 1772), *Balfaldera* (> 1858), *Bandbulle* (> 1810), *barbarisere* (> 1843), *Batteri 3* (spøg. om en samling flasker, > 1864), *Bavian* (som skældsord, > 1787), *bedugget* (beruset, > 1779), *begrave 3.2 slutn.* (m.h.t. lovforslag olgn., > 1858), *Ben 2.3* (om bierhverv, > 1858).

Tilbagedateringer fra ODS-belæg ældre end ordbogens samtid: *afgaa 2.1*, få ell. tage sin afsked, VSO.1793 > HolbO., *afskedige 2*, opsige fra et arbejde, (< vAph.(1764)) VSO.1793 > HolbO., *Allongeparyk* 1869 > 1740, *amusant* 1800 > 1751, *analysere* ca. 1780 > HolbO., *Ballastsand* 1860 > 1788, *Balsam*, i bet. lindring, trøst, ca. 1830 > 1732, *III.bar 3.2*, fx som bare pokker, ca. 1840 > 1785, *Barberkost* MO.1859 > 1797, *give barnet navn* (sp.1160,2), sige noget rent ud, ca. 1800 > 1701, *barrikadere* JBaden.FrO.1820

> 1765, *Bassin* 1808 > 1765, *II.basta*, interj., 1807 > 1749, *befuldmægtige* VSO.1793 > HolbO., *Benrad*, om døden, 1824 > 1776, *Berøringspunkt* 1843 (billedl.) > 1799 (egl.); jf. ordet *Berøringsflade* S, 1805, den billedl. brug belagt ca. 1935, *betalbar* 1873 > 1818, *Beton* (< MO.1859) 1873 > 1868, i sammensætningen *Betongrus* 1838.

En tilbagedatering med ret få år har især interesse i nærheden af ordets eller ordanvendelsens formodede fremkomst. Nogle eksempler: *alkoholisere*, mht person, 1908 > 1889, *Alkoholist* 1917 > en fremmedordbog 1899; ordet er for øvrigt trængt af *Alkoholiker* S, 1905, *Arkæologi* 1815 > 1807, *Automatkafé* 1912 > 1902, *Bajer*, bajersk øl, 1877 > 1857; ordet har udkonkurreret substantiveringen *bajersk*, (fx "sit Kruus Bajersk") som er brugt i 1850'erne, *Ballon*, i bet. haløj, ballade, 1870'erne > 1859, *Banjo* 1894 > 1883, *Barselheber* ca. 1850 > 1827, *Bersærkergang* beg. af 1800-tallet > ca. 1780.

Det modsatte af tilbagedatering kan vel kaldes fremdatering, det vil sige en angivelse af yngre (senere) belæg end hidtil kendt, og en sådan fremdatering må ifølge sagens natur angå ordanvendelser som ikke mere er i brug, ellers var der jo ikke tale om belæg yngre end hidtil kendt.

Ordanvendelser der i ODS er betegnet med kors har som sidste belæg et citat fra det 18. eller 19. århundrede. Hvis der nu kan suppleres med et belæg der er betydelig yngre, er ordet fremdateret: *Beliggende*, beliggenhed: sidstebelæg H.C. Andersen > 1895, *Be-slutning 1*, ende, afslutning: sidstebelæg før midten af 19. årh. > 1886, *bevant 1*, øvet, erfaren: sidstebelæg 1816 > 1932 (en bevant murersvend; i en stillingsannonce, som eventuelt er dialektpræget).

Et yngre citat i ODS-S er ikke altid en fremdatering, ofte er det blot et yngre eksempel på noget der stadig er fuldt brugeligt. Det er medtaget fordi ordbogsværkets nutid bør være repræsenteret i kildematerialet (men ikke nødvendigvis slutningen af ODS-perioden). Eksempler på citatmangel af denne art i ODS: *alenlang*, overf.: yngste belæg i ODS er fra midten af 19. årh., + yngre: 1955, *almindeligvis*: H.C. Andersen yngst, + yngre: 1953, *Angstskrig*: kun ordbøgerne VSO.1793 og MO.1859, + yngre: 1954.

Ved de egentlige fremdateringer er det nødvendigt med en brugsrestriktion. Det kan i ODS-S hedde: "+ yngre (men aftagende brug i 20. årh.);" eller: "+ yngre (men ikke alm. i 20. årh.)."

En vis del af det ordstof der var almindelig brugt da begyndelsen af ODS blev redigeret, er i slutningen af ODS-perioden godt på vej ud af sproget. Dette forhold kan give anledning til bemærkningen "aftagende brug i 20. årh.;" et påfaldende sent citat medtages gerne, for at vise ordenes sejlivedhed. Eksempler på ord med aftagende brug i 20. årh. er: *agtbar* (som får tilføjet et Karen Blixen-citat), *Arbejdskone*, *-kvinde*, *Automobilist* (kan ses i Rigsdagstidende 1946), *basune* (hvorimod *basunere*, der ifølge ODS var mindre brugt, det var måske ret nyt omkring 1900, får bemærkningen "øget brug i 20. årh.;"

Nu kan man spørge om dateringen i ODS-S, også dateringen af de ord som er nye i forhold til ODS, er til at stole på. - Det er den naturligvis ikke (jævnfør de ndf. nævnte pro-

blemer vedrørende førstebelæg og sidstebelæg). Mange af ordene må kunne findes 10-20 år tidligere, men næppe mange af de daterbare ord er 50-100 år ældre end vort ældste belæg (når vi ser bort fra tiden før 1700, som vi ikke ved nok om og som ligger uden for ordbogens periode). Og med hensyn til sidstebelæg kan mange af de ord der er angivet som kun tilhørende ældre sprog, dvs sproget før ca. 1900, utvivlsomt påtræffes senere end vi har kunnet vise.

Datering af etymologiske eller semantiske enheder

Spørgsmålet om hvornår ordene dukker op i et bestemt sprog kan siges at høre med til de etymologiske oplysninger. Elof Hellquist medtager i sin Svensk etymologisk ordbok, hvis 1. udgave er fra 1922, årstal for ordenes tidligste optræden i svensk, så vidt det har været muligt. I Niels Åge Nielsen, Dansk Etymologisk Ordbog, 1966 (og senere udgaver) er ordets alder i dansk ikke anført direkte, men der er givet oplysning om forekomster i runedansk, gammeldansk eller ældre nydansk.

Hvis man i en etymologisk ordbog vil se hvornår de ord vi bruger i dag er kommet ind i sproget, så bliver svaret måske misvisende. For det man ønsker at vide noget om er de kendte ordanvendelser, altså ordet i bestemte semantiske anvendelser, snarere end det er den etymologiske ordform. En sådan semantisk ordanvendelse kan være en ordbetydning, en enhed af flere ord (idiom, term) eller en produktiv orddel (affiks, sammensætningsled).

De følgende eksempler skal vise at en rent etymologisk datering ofte rammer ved siden af det moderne ords hovedbetydning (jeg følger i eksemplerne ODS-artiklernes struktur og supplerer med ODS-S-artiklernes oplysninger):

Ambulance 1 feltlazaret ell. -sygevoan. 1837ff. +2 civil sygevoan. 1861ff. +3 billedl. 1932ff.

annoncere 1 bekendtgøre. 1700ff. 2 bekendtgøre gennem annoncer. 1825ff.

annektere 1 tilknytte. ca.1700ff. 2 om stat: sætte sig i besiddelse af et landområde. 1870ff.

anløbe 1 bevæge sig mod et mål. 1.1 nærme sig i løb ell. med fart. VSO.(1793), ikke brugt o. 1900. 1.2 sejle ind til (en havn olgn.) for et kortere ophold. 1807ff. +|| (fagl.) om andet end skibe. 1934ff. +1.3 anfalde; angribe. ældre nydansk og 1838. 2 om visse overfladeforandringer (ved metaller). 2.1 (især tekn.) intr.: miste den oprindelige farve og glans. 1777ff. 2.2 (tekn.) trans.: fremkalde en farveforandring på (et metals overflade). 1802ff. 2.3 part. *anløben* brugt som adj. om metaller: 1700ff. || overtrukket med skimmel olgn. om vin: Moth.(1700), ikke brugt o. 1900. || (især naturv.) svagt farvet. 1902ff. || (ofte spøg.) om virkninger af alkoholnydelse. om næsens farve: slutn. af 19. årh. let beruset: 1880ff. || ikke ren ell. frisk; angrebet; let fordærvet. fagl., fx om træ: 1900ff. om noget i moralsk henseende let fordærvet: 1880ff.

Det etymologiske ord *anløbe* er altså kendt i ældre nydansk i betydningen 'angribe' (i Dansk Etymologisk Ordbog er der ikke betydningsangivelse). I almensproget i dag bruges ordet især om skibe der sejler ind til en havn; participiet *anløben* eller *anløbet* har skilt

sig ud og opfattes nu som et selvstændigt ord. Eksemplet skal illustrere at man ikke kan "løbe an på", at den etymologiske ordform i sin ældste anvendelse ligner de moderne betydninger.

Eksempler på enheder af flere ord: Ordforbindelsen (idiomet) *løbe an på*, med betydningen 'regne med at noget vil ske eller forholder sig på en vis måde', er måske samme etymologiske enhed som anløbe, men jo en selvstændig semantisk enhed; under II.løbe 22.3 i ODS er det ældste belæg fra ca. 1880. Idiomerne *i lange baner* og *i brede baner*, begge med betydningen 'i mængde', var ifølge ODS almindelige i daglig tale o. 1900; det sidste udtryk har ikke holdt sig i sproget. Faglige termer som *betinget dom*, 1911ff., og *+ betinget refleks*, 1932ff., bør naturligvis dateres som selvstændige enheder.

Produktive orddele med en særbetydning er fx: *-apostel* i ord som afholdsapostel, sandhedsapostel, sundhedsapostel, der er kendt fra 1911 (afholdsapostel); *-bank* om et depot af andet end penge er kendt i dansk siden 1943, ældst i blodbank. Også disse ord-dannelsesmuligheder er det rimeligt at få dateret.

Problemer vedrørende førstebelæg og sidstebelæg

En sikker datering forudsætter naturligvis en omfattende registrering af ordforrådet gennem tiderne. En forbedret registrering af dansk vil indtil videre medføre tilbagedateringer og fremdateringer. Men der er en del særlige forhold at betænke ved efterforskningen af ordenes liv og død.

En vurdering af ordets art:

Visse sider af ordforrådet lader sig vanskeligt datere, eller sagt på en anden måde: det foreliggende førstebelæg vil næppe ligge tæt på ordanvendelsens oprindelse. Dette gælder

- a. De ældste ord i sproget, arveordene, som går længere tilbage end kilderne rækker.
- b. De udprægede talesprogsord fra både ældre og nyere tid. En stor del af dialekternes ordforråd er kun kendt fra optegnelser inden for de sidste 100 år, men er naturligvis ældre. Slangudtryk kan leve kort eller de kan leve længe, det er svært at vide, og kil-dematerialet svigter.
- c. De selvfølgelige eller tilfældige orddannelser som står til rådighed i sproget. Et ord som *Beslutningskraft* anså ODS-redaktøren for dødt, modsat beslutningsdygtighed og -evne; ODS-S viser et 1950-citat, og ordet kan sandsynligvis også træffes i begyndelsen af århundredet; det er ikke almindelig brugt, men kan nydannes af den der vil.
- d. Visse låneord. Hellquist gør i sin indledning opmærksom på at han har undladt at datere en del romanske eller klassiske låneord, da deres ældste optræden ofte er rent tilfældig, som citatord, og måske ikke har direkte historisk sammenhæng med den senere, faste brug.

Den anden yderlighed mht dateringsmuligheder er de kulturord som kan dateres helt præcist. Et kendt eksempel på et ord med fødselsattest er ordet *bil*, der blev foreslået til

en dagbladskonkurrence i 1901. Inden for videnskab og teknik er der en række internationale ord som er dannet og defineret på kendte tidspunkter af bestemte ophavsmænd; også visse officielle ord vedrørende samfundsforhold er dannet eller indlånt på bestemte, kendte tidspunkter.

En vurdering af dateringskilderne:

Det bedste kildemateriale er naturligvis citater fra løbende tekst hvor tid, sted, situation og sprogbruger er kendte størrelser. Ordbøger kan være farlige som dateringskilde, da ordbogsord ikke altid afspejler den faktiske sprogbrug. Oversættelsesordbøger har ofte konstrueret ord ud fra det fremmede ordstof³. Fremmedordbøger, især de ældre, støtter sig på udenlandske ordbøger og har ikke altid været så nøjeregnende med om ordene var at finde i danske tekster.

Øjenvidneberetninger er nyttige. ODS citerer ofte de ældre ordbøger for bemærkninger som "nyt ord" (se fx u. betone) eller bemærkninger om at et ord er ved at udgå. ODS selv er et øjenvidne til sproget i redaktionstiden. I 1955 fik vi Dansk Sprognævn, som bl.a. har til opgave at holde øje med sprogforandringerne. På grundlag af sprognævnets iagttagelser har vi fået ordbogen Nye ord i dansk 1955-75 (1984), redigeret af Pia Riber Petersen (nu: Pia Jarvad). En registrering af de ord som mange vil anse for døde, men som alligevel jævnlgt dukker op, har derimod ikke interesseret sprognævnet ret meget.

En vurdering af frekvens og stilværdi:

Begrebet datering rejser spørgsmålet om kriterier for hvornår en ordanvendelse kan siges at tilhøre sproget. Nye ord i dansk 1955-75 indeholder ifølge indledningen (s. 11) "nye ord og vendinger som er blevet almindelige i perioden 1955-75"; det er altså ikke nødvendigvis absolutte førstedateringer. Hertil svarer forekomsten af oplysninger som: "nu i almensproget" (u. socialgruppe), "øget brug" (u. ulækker), "øget brug i almensproget" (u. risikogruppe, snalret). Ord der fx kun er kendt inden for dialekterne tilhører naturligvis dansk, men dog ikke det fællesdanske sprog. Det samme gælder ord fra det interne fagsprog.

Ord kan komme ind i sproget pludseligt, og ord kan siges at komme langsomt; denne skelnen forudsætter at man tager hensyn til frekvens og stilværdi. Nogle ord kan leve et kort og hektisk liv som modeord.

Ord kan gennem en meget lang periode være på vej ud af sproget; måske bliver de stående i en niche og fungerer som højtidelige eller som uhøjtidelige, spøgende ord, eller de giver teksten et gammeldags præg.

Dateringsforskning er et område hvor der endnu ikke er sket ret meget. En sådan forskning kan have interesse for både sprog-, kultur- og litteraturhistorien.

Noter

1. Robert and Clarence Barnhart: *The Dictionary of Neologisms*, i: Wörterbücher, Dictionaries, Dictionnaires. Red. af F.J. Hausmann et al. Berlin & New York. II. 1990. s. 1159-66. - Oscar Reichmann: *Wörterbücher archaischer und untergegangener Wörter*. smst. s. 1153-58. - Wilhelm Kesselring: *Das chronologische Wörterbuch*. smst. s. 1342-48.
2. Følgende kildeforkortelser forekommer i eksemplerne: *HolbO.* = Aage Hansen og Sv. Eegholm-Pedersen: *Holberg-Ordbog*. I-V. 1981-88. - *JBaden.FrO.* = Jacob Baden: *Ordbog over de af fremmede Sprog laante Ord*. 2. udg. 1820. - *MO.* = C. Molbech: *Dansk Ordbog*. 2. udg. I-II. 1859. - *S* = der henvises til en artikel i ODS-S. - *S&B.* = Th. Sundbye & E. Bartiel: *Dansk-norsk - fransk Haand-Ordbog*. I-II. 1883-84. - *vAph.(1764)*. = H. von Aphelen: *Kgl. Dansk Ord-Bog*. I. Dansk og Tydsk. 1764. - *VSO.* = *Dansk Ordbog* udg. under Videnskabernes Selskabs Bestyrelse. I-VIII. 1793-1905.
3. Bo Ralph: *The older dictionaries as sources for Nordic language history*, i: J. Louis-Jensen & J.H.W. Poulsen (eds.): *The Nordic Languages and Modern Linguistics* 7. Tórshavn 1992 s. 493-509.

Genbrug af Ordbogsdata: En dansk Ordbank

Dorthe Duncker og Ole Norling-Christensen

Den Danske Ordbog (DDO) i 6 bind skal udkomme 1998-99. Den skal beskrive sproget efter ca. 1950 med hovedvægt på perioden 1983-92, ordbogens såkaldte *primærperiode*.

Datamatiske hjælpemidler benyttes i stort omfang som støtte for det redaktionelle arbejde, og vi har samlet en væsentlig del af ordbogens kilder i to store databaser. Den ene er DDO's tekstkorpus på 40 mio. ord, der dækker primærperioden (Kristensen 1994). Den anden er *Ordbanken*, der omfatter bearbejdede data fra en række eksisterende ordbøger, hvis oplysninger er af betydning for DDO. Det er Ordbanken, og vekselvirkningen mellem denne og korpuset, som er emnet for denne artikel.

Tanken om at samle og koordinere de leksikalske oplysninger om dansk, som allerede foreligger, men er spredt på mange forskellige kilder, er ikke ny. Den var udgangspunkt for det forskningssamarbejde, som indledtes i 1982 og siden fik navnet DANLEXgruppen (DANLEX 1987a); den vandt tilslutning såvel i den danske ordbogsverden (DANLEX 1987b), som i bredere kredse af professionelle ordbrugere (DANLEX 1989); og den var en selvfølgelig og nødvendig del af den plan (Dansk Ordbog 1989), som blev grundlaget for DDO-projektet med dets redaktionsperiode på kun 6 1/2 år.

Mens sproget i primærperioden i det væsentlige skal beskrives på grundlag af korpuset, er hovedkilderne til tiden før 1983 Dansk-fransk Ordbog (B&H) og Dansk-engelsk Ordbog (V&B), hvis førsteudgaver begge byggede på Ordbog over det Danske Sprog 1-28 (ODS). Sammen med Retskrivningsordbogen (RO), den officielle danske norm for stavning og bøjning, samt Dansk Sprognævns Ordregister (DSO) (nøglen til en samling på henvend en million excerpter, som Sprognævnet har indsamlet siden sin oprettelse i 1955), har disse ordbøger gennemgået en omfattende datamatisk behandling, der har resulteret i en database over mere end 300.000 danske ord med angivelse af ordklasse, ortografiske varianter, bøjnings- og orddannelsesoplysninger, betydningsdifferentiering og eksempler på mulig kontekst.

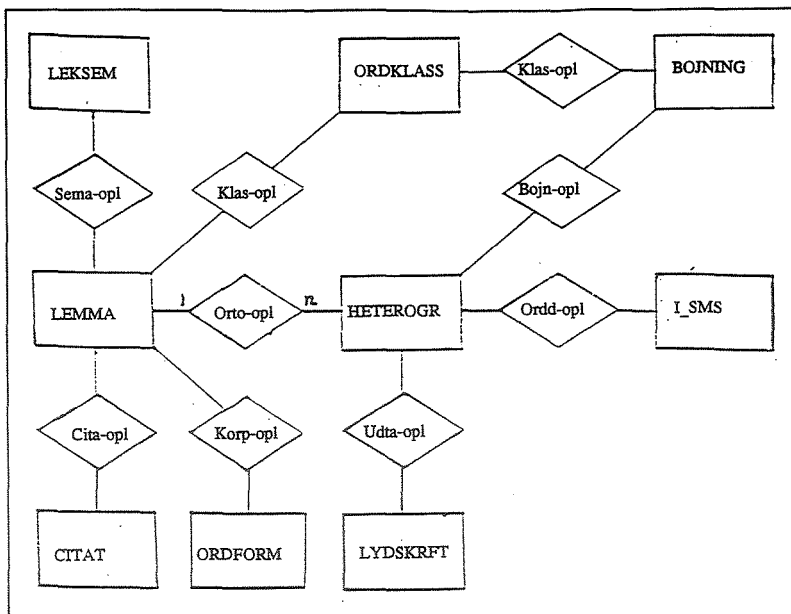
Udover funktionen som kilde til ord, betydning og brug før 1983, og til korrekt dansk stavning og bøjning, har Ordbanken endnu et væsentligt formål, nemlig at identificere samtlige de ordformer, som forekommer i korpus, dvs. henføre dem til en eller flere bøjningsformer af et eller flere lemmer. Dette er nemlig en (nødvendig, men ikke tilstrækkelig) forudsætning for at kunne søge præcist i korpus efter alle former af et givet lemma. Desuden giver det et vigtigt bidrag til den ordbogsmæssige beskrivelse af, hvordan den faktiske sprogbrug forholder sig til normen.

Materialet fra ordbøgerne bliver derfor suppleret med de "nye ord", som findes i korpus, og med de yderligere stave- og bøjningsoplysninger som kan udledes af korpus. Si-

den hen vil også udtaleoplysninger blive tilføjet, samt formentlig oplysninger om (hel eller delvis) synonymi.

Ordbankens struktur og indhold

Ordbankens opbygning fremgår af entitet-relationsdiagrammet i fig. 1 (jf. fx DANLEX 1987a: 234-239). Hvert rektangel repræsenterer en entitet og hver rombe en relation mellem to entiteter. Det anvendte databasesystem er Paradox (version 4.0; Borland International Inc.). Med få undtagelser er hver entitet og hver relation realiseret som en særskilt tabel, hvortil der er knyttet et eller flere indekser. Den typiske entitetstabel rummer kun to felter, nemlig en entydig identifikation af posten i form af et nummer og et indholdsfelt. En relation sammenknytter to bestemte entiteter, eller evt. én entitet med sig selv. En post i en relation rummer derfor først og fremmest felter med reference til nummeret på en bestemt post i hver af de entiteter, som relationen knytter sammen. Hertil kan komme et eller flere attributter, felter som nærmere bestemmer arten af den pågældende relation. Ordbankens relationer indeholder foreløbig ét attribut, nemlig feltet *Kilde*, som redegør for, hvilken ordbog eller anden kilde den oplysning, relationen udtrykker, stammer fra. Denne måde at strukturere databasen på (den er, i databaseterminologi, fuldt normaliseret) gør det let at tilføje nye relationer mellem ordene, fx synonymi, etymologisk slægtskab, eller referencer mellem sammensatte ord og de led, de er sammensat af.



Figur 1. Ordbankens opbygning af *entiteter* (rektanglerne) og *relationer* (romberne). De særne navne skyldes reglerne for fil-navngivning: højst 8 tegn; ingen danske særbogstaver

Entiteterne

Centralt står LEMMA, den enhed ordbogsartiklen skal beskrive. Det er en abstrakt enhed, som ortografisk i en del tilfælde kan realiseres ved forskellige (grund)former, her kaldt HETEROGRAfer. Til LEMMAet er desuden knyttet CITATer, LEKSEMer, ORDFORMer og en ORDKLASSe.

Til hver HETEROGRAf er knyttet: en BOJNING (oplysning om bøjningsmønster, hvorudfra alle former kan genereres), en eller flere orddannelsesoplysninger (I_SMS), der angiver i hvilken form 'heterografen' indgår som førsteled i sammensætninger; hertil vil komme en eller flere UDTALer, dvs. fonetisk transskription af grundformen.

CITATerne er de danske udtryk fra V&B og B&H.

LEKSEMerne er først og fremmest V&B/B&H's betydningsadskillende oplysninger; hvis der er anført flere betydninger, medtages en eventuel nulmarkeret (første) betydning også.

FORMerne er fortegnelsen over samtlige forekommende ordformer i korpus med reference til de LEMMAer, som formerne iht. rent morfologiske kriterier kan henføres til.

ORDKLASS er blot en nummereret liste over ordklassernes navne; relationen Klas-opl, som knytter den til LEMMA og BOJNING er ingen tabel, men blot udtryk for, at ordklassen i disse to tabeller angives ved dens nummer i ORDKLASS-tabellen.

De fleste af relationerne bliver *mange til mange relationer* (n-m); for overskuelighedens skyld fastholdes dog lemma-heterograf relationen som en 1-n relation; det indebærer fx, at selv om "spise vb." og "spise sb." har samme grundform, har de hver sin post i heterograf-tabellen.

Fra ordbøger til Ordbank

Ordbøgerne, som udgør Ordbankens hovedkilder, foreligger alle i maskinlæsbar form, men er ikke som udgangspunkt sammenlignelige. Hver kilde har derfor gennemgået en forbehandling, hvorunder samtlige relevante oplysningstyper er blevet bestemt og omsat til et fælles format. Vi har i denne forbindelse delt artikler, som indeholdt mere end ét LEMMA, fx "svejs^e, -ede, i teknisk sprog også -te." (RO) og "øl -let, øl -lene; om flaske e.l. med øl: øllen, øl el. øller, bf. flt. ølle(r)ne." (RO).

Den første og centrale kilde i opbygningen af Ordbanken er RO. Derefter følger DSO, B&H og V&B. RO fungerer som bankens ortografiske lovtekst. Det vil sige, at hvis vi finder det samme opslagsord med samme oplysninger om ordklasse og køn, men med forskellige bøjningsoplysninger i en senere kilde, så sejrer foreløbig RO's bøjningsoplysninger. Fx angiver RO, B&H og V&B hver sin bøjningsoplysning ved *krysantemum* ("-(m)en, -(m)er el. krysantemum el. krysantemer, bf. flt. krysantemum(m)e(r)ne el. krysantemerne." (RO); "(c) (- el. -mer)" (B&H); "en -mum(m)er el. pl. d.s. el. -mer" (V&B)), men kun oplysningerne fra RO bliver indført i banken.

RO indeholder to typer opslagsord: dem med bøjningsoplysninger og dem med lodret

streg. Den lodrette streg markerer, at bøjningsoplysninger kan søges under sidsteleddet som opslagsord. Disse to typer kan imidlertid kombineres, således at et opslagsord både har lodret streg og bøjningsoplysninger. Om denne type siger RO: "Oplysninger om bøjning mv. er kun anført, hvis de afviger fra oplysningerne under sidsteleddet som opslagsord, eller hvis der er mere end ét opslagsord med samme stavemåde som sidsteleddet" (RO: 16). Om det er den ene eller anden årsag, der har motiveret oplysningerne, kan man erfare ved at slå op under sidsteleddet. Ved fx *kobber/stik*, der har anført bøjningen "-ket, kobberstik -kene", kan man slå op under *stik* og finde to homogرافnumre: "1. *stik* -ket, stik -kene" og "2. *stik* ubøjel. to.", mens man ved fx *kobber/stikning* med bøjningsoplysningen "-en" kan slå op under *stikning* og se, at oplysningen hér afviger fra sammensætningens: "-en, -er".

Opslagsord uden lodret streg lagres hele i Ordbanken med deres oplysninger. Op-slagsord med lodret streg, men uden bøjningsoplysninger, lagres ved at oplysningerne under sidsteleddet som opslagsord kopieres; fx *kogelflæsk*, *kogelgrej* og *kogeljomfru*, der forsynes med bøjningsoplysningerne fra hhv. *flæsk*, *grej* og *jomfru* som opslagsord. Op-slagsord, som både har lodret streg og bøjningsoplysninger, lagres med de anførte op-lysningsord. I de tilfælde, hvor sidsteleddet ikke er en homogراف, registreres det desuden, at det aktuelle sidsteled kan ændre bøjning i forbindelse med sammensætning. Fx angiver RO ved *hår* som opslagsord bøjningen "-et, hår -ene". Blandt sammensætningerne med *-hår* som sidsteled findes *fimrelhår*, *knurlhår* og *øjenhår*, hvis bøjning kan søges un-der *-hår* som opslagsord. Man møder imidlertid også sammensætninger med *-hår*, hvor bøjningen afviger fra *hår* som opslagsord, fx *afrolhår*, *pagelhår* og *plyslhår*, som har bøj-ningsoplysningen "-et". Denne type oplysning - at *-hår* i sammensætninger med betyd-ningen "frisure" mister flertal - kan anvendes, når bøjningen for ord fra de følgende kil-der skal bestemmes.

Efter RO er den næste kilde, som behandles, DSO. Strukturelt kan ordene fra DSO sammenlignes med RO's opslagsord med lodret streg, og de lagres i Ordbanken efter samme principper som disse. Hvis sidsteleddet er en homogراف, kan bestemmelsen ikke foregå fuldautomatisk, men må støttes ved manuel stillingtagen i det enkelte tilfælde. Når dette arbejde er tilendebragt, og både RO og DSO er færdigbehandlet, indeholder Ordbanken godt 230.000 forskellige lemmaer.

Den tredje kilde, som inddrages, er den danske side af B&H. B&H behandles før V&B, fordi B&H anfører ordklasse og køn ved sammensætninger, hvilket V&B ikke gør konsekvent. Disse oplysninger er væsentlige, da de kan spare en del håndarbejde i for-bindelse med entydiggørelse af homografe sidsteled. Fx vil *hovednøgle* på denne måde kunne behandles fuldautomatisk, idet B&H anfører ordet med sammensætningsgrænse *hovednøgle* og køn, således at det er muligt fuldautomatisk at vælge mellem de to homo-grafnumre i RO "1. *nøgle* -n, -r (til lås)" og "2. *nøgle* -t, -r (garnnøgle)" og tilknytte oplys-ningen fra "1. *nøgle*" til *hovednøgle*. I V&B findes *hovednøgle* med sammensætningsgræn-se, men uden oplysning om ordklasse og køn. Da ordet allerede er bestemt i den forrige kilde og ikke er en homogراف, kan det uden videre genkendes automatisk.

Umiddelbart kan det forekomme overdrevet, at vi gør os så store anstrengelser for at genkende et ord, vi allerede har lagret i banken. Det skyldes to forhold: For det første er det nødvendigt at kunne bestemme de enkelte opslagsord i de forskellige ordbogskilder entydigt, således at vi ikke kommer til at lagre det samme ord flere gange, bare fordi vi ikke har været i stand til at tolke kildernes bøjningsoplysninger korrekt. For det andet ønsker vi også at genbruge og lagre artiklernes betydningsspecifikationer (LEKSEMerne) og danske vendinger og udtryk (CITATerne) fra B&H og V&B. Disse oplysninger skal bruges under opbygningen af artikelskeletter, der i første omgang skal fungere som støtte for redaktionsprocessen. Det er derfor vigtigt, at vi med sikkerhed kan knytte disse oplysninger til det rette LEMMA i Ordbanken.

Opslagsordene fra B&H og V&B lagres i øvrigt efter samme principper, som gjaldt for RO's ord med lodret streg. Det vil sige, at nye sammensætninger lagres, hvis sidsteleddet allerede kendes. Giver ordbogen selv oplysninger om bøjning ved (usammensatte) opslagsord, som ikke kan genkendes på grundlag af Ordbanken, lagres ordet med disse bøjningsoplysninger.

Supplering fra korpus

På dette tidspunkt rummer Ordbanken alle genkendte, forskellige opslagsord fra de fire ordbogskilder og er nu klar til at blive konfronteret med korpuset. I første omgang vil vi gerne vide, hvilke korpusord der er kendte FORMer af kendte LEMMAer. Det undersøges, om det enkelte korpusord findes blandt de former, der kan genereres af HETEROGRAfernes bøjningsoplysninger (BOJNING). I så fald lagres korpusordet som FORM i relation til det eller de LEMMAer, den eller de aktuelle HETEROGRAfer tilhører.

For at kunne foretage denne bestemmelse må vi kunne fortolke lemmaernes bøjningsoplysninger, og hér støder vi på et problem, som vi har arvet primært fra RO. Langt den overvejende del af disse oplysninger stammer fra RO, dog suppleret med oplysninger fra B&H og V&B. Bøjningsoplysningerne er ikke angivet udtømmende, men peger hver især på et bøjningsmønster.

RO er skrevet med danske sprogbrugere for øje, men ikke alle danske sprogbrugere. I forbindelse med præsentationen af den nye Retskrivningsordbog i 1986 blev det sagt, at det "er - og har altid været - en stiltiende forudsætning at de officielle retskrivningsordbøger med tilhørende retskrivningsvejledninger ikke sigter mod de danske skriftsprogsbrugere i bred almindelighed, men mod skriftsprogsbrugere over et vist niveau. Groft sagt er Retskrivningsordbogen først og fremmest beregnet på dem der i forvejen sidder inde med en høj grad af skriftsprogsbeherskelse, og de emner der behandles i Retskrivningsreglerne er først og fremmest dem hvor 'ordentlige sprogbrugere' kan tillade sig at være i tvivl" (Galberg Jacobsen 1987: 33).

Nu er *brugeren* i denne sammenhæng ikke noget menneske, men derimod en række computerprogrammer. Disse programmer indeholder ikke fra første færd den viden, man altså forudsætter af brugerne. Når man automatisk skal tolke RO som udsagn om

normen, rejser der sig derfor det konkrete spørgsmål: Hvad er det disse skriftsprogsbrugere véd, og som ikke står i RO, men som man skal implementere i programmerne for at sikre, at de kan administrere normen forsvarligt?

Opgaven, vi står overfor, er altså at *efterkonstruere* den kompetence, som disse sprogbrugere forventes at sidde inde med. Ellers kan vi ikke få et program til at afgøre, om et vilkårligt tekstord fra korpuset falder inden eller uden for normens rammer. Og kan vi ikke det, kan vi heller ikke leve op til målsætningen om at bruge Ordbanken til at sammenligne norm og usus.

De oplysninger, RO giver om et ords bøjning, peger på et bøjningsmønster, hvor sprogbrugeren - eller her programmet - selv må fylde hullerne ud. Men et program kan ikke bringes til at se, om nogle af de herved dannede former er umulige. Generelt er det fx således, at der ved substantiver er "givet oplysning om bestemt form ental og ubestemt form flertal, når disse kendes. Der er kun givet oplysning om bestemt form flertal, når den ikke dannes ved tilføjelse af endelsen *-ne* til den ubestemte flertalsform" (RO: 18). Ved "*tilfælde* -t, flt. tilfælde" er der ikke oplyst om bestemt form flertal, som derfor kan dannes ved at tilføje *-ne* til *tilfælde*: *tilfældene*. Ved "*gentleman*, en, flt. gentlemen" er formen ligeledes ikke anført og skulle således kunne dannes efter samme princip, hvilket de fleste danske sprogbrugere nok vil undlade.

En holdning til denne type overgenerering kunne være at sige, at hvis en dannet form er umulig, vil den næppe heller forekomme i korpuset og dermed ikke forårsage støj i form af utilsigtede bestemmelser af korpusord. Imidlertid peger forholdet på en mangel ved denne traditionelle måde at oplyse om et ords bøjning på. En udeladt oplysning om en bøjningsform kan betyde ét af to: enten at formen ikke findes, eller at det er forudsat viden, om den dannes. En sprogbruger "der i forvejen sidder inde med en høj grad af skriftsprogsbeherskelse" vil måske nok kunne skelne, men det kan et program ikke.

Som udgangspunkt har vi valgt at danne netop de former, som bøjningsoplysningerne giver os grund til at antage er mulige. Herunder inddrager vi også de retningslinjer som ellers findes i RO (RO: 13-23; 549-571). Desværre medfører denne normlydige adfærd, at vi kommer til at undergenerere, hvilket direkte har konsekvenser for redaktionsarbejdet. Problemet med undergenereringen er i praksis, at søgninger i korpuset bliver ufuldstændige og dermed upålidelige. Hvis man fx søger i korpuset efter forekomster af adjektivet *smuk* og alene følger RO's mønster "*smuk* to., -t, -ke", finder man belæg på formerne *smuk*, *smukt*, *smukke* og *smukkes*. Man finder ikke i samme søgning *smukkere*, *smukkest* og *smukkeste*, som også er belagt. Vi kan altså ikke beskrive korpuset tilfredsstillende, hvis vi nøjes med at holde os inden for normens eksplicit angivne rammer.

Undergenereringen rammer især - som antydnet - én gruppe af former, nemlig adjektivernes gradbøjning. RO oplyser kun om gradbøjning ved enkelte adjektiver, nemlig "hvor der sker vokalskifte eller ændring til en anden stamme, og ved tillægsord, der ender på *-som*" (RO: 20). Adjektiver, som ikke opfylder disse krav, men som alligevel kan gradbøjes, savner oplysning herom. Det er sagtens muligt at implementere reglerne for, hvordan formerne dannes i et program, men programmet vil i så fald gradbøje alle de adjektiver, det kommer i nærheden af, og ikke kun dem, der faktisk kan gradbøjes.

Når vi ikke i Ordbanken har tilstrækkelig viden om gradbøjningen til generelt at danne formerne, må vi lede efter dem i korpuset. Det gør vi ved hjælp af morfologisk analyse, hvor vi genkender et korpusord som enten komparativ eller superlativ, hvorefter oplysningen om gradbøjning indføres i banken ved det aktuelle adjektiv. Herved får vi i det mindste viden om, hvilke adjektiver sprogbrugerne gradbøjer.

De tekstord, vi ikke har kunnet bestemme på grundlag af bankens aktuelle oplysninger, udgøres primært af sammensætninger, afledninger og bøjningsformer heraf. Hvis vi kan opløse dem, vha. af de orddele vi allerede kender, indføres de i banken. Dette er for sammensætningernes vedkommende ikke helt så enkelt, som det lyder. For det første gør de samme problemer med homografe sidsteled sig gældende, som de gjorde i ordbogskilderne. Vi har altså stadig brug for manuelt at entydiggøre disse sammensætninger. For det andet kan en del sammensætninger tilskrives mere end én analysestruktur. I tilfælde af flere strukturer er det altså nødvendigt med en manuel indgriben for at vælge blandt resultaterne. Fx vil en dansk sprogbruger nok ikke spekulere over, at korpusordet *reklamebudskaber* kan tolkes som andet end ubestemt form flertal af et *reklamebudskab*. For et program, der intet kender til virkelighedens verden, kan ordet imidlertid også tolkes som en *reklamebudskaber* (jf. fx Brandt 1993).

Nu står vi tilbage med en samling korpusord, der hverken er genkendelige bøjningsformer eller sammensætninger og/eller afledninger. Som man kunne vente, indeholder restmængden en del regulære slå- og stavfejl, der er også ord fra andre sprog, ad hoc-dannede interjektioner og nogle helt nye lemmaer, men hvad mere interessant er, så rummer listen også en del *regelmæssige variationer* over de bøjningsoplysninger, vi allerede har i banken samt en del supplerende former. Fx kan flertalsformer af substantiver som if. RO ikke kendes i flertal dukke op hér, fx *polterabender* ("*polterabend*, en." (RO)) og *sugeevner* ("*sugeevne* -n." (RO)). I så fald opdateres Ordbankens bøjningsoplysninger (BOJNING) med oplysning om, at denne brug forekommer i korpuset.

Endelig finder vi her eksempler på afvigelser, som kolliderer med autoriserede former, men som er dannet regelmæssigt, det vil sige dannet ved hjælp af kendte fleksiver og kendte morfografematiske ændringer, fx *brudet* ("*3. brud* -det, brud -dene." (RO)), *sweatre* ("*sweater* -en, -e sweaterne." (RO)) og *karnevallet* ("*karneval* -et, -er" (RO)). Disse afvigelser indføres også i banken, fordi de spiller en rolle, når et helt lemma skal findes i én søgning, men det oplyses at formen står i opposition til normen. Oplysninger om regelmæssig afvigelse er væsentlige, når vi skal undersøge, hvordan sprogbrugerne bruger ordene i virkeligheden.

Anvendelser

Oplysningerne om ortografiske varianter og bøjningsparadigmer bruges af Den Danske Ordbogs konkordansprogrammel, hvorved søgning på én grundform automatisk kan udvides til søgning efter samtlige stovevarianter i alle bøjningsformer (autoriserede eller ej). Desuden bruges bøjningsoplysningerne til første trin af en automatisk homografseparation (ordklassetagging af korpus), og til undersøgelse af, hvilke bøjningsformer af

ordene der faktisk forekommer i korpus; herved fås et empirisk grundlag for Den Danske Ordbogs behandling af bøjningsoplysninger. Ud fra databasen konstrueres endelig de "artikelskeletter", som ordbogsredaktørerne kan benytte som udgangspunkt for redigering af Den Danske Ordbogs artikler. Ingen oplysning fra Ordbanken bliver altså viderebragt i ordbogen, uden at den har været vurderet af mennesker.

Referencer

- B&H: Blinkenberg, Andreas og Poul Høybye: *Dansk-fransk ordbog / Dictionnaire Danais-Français*. 4. udg. ved Jens Rasmussen & al. Bind 1-2. København 1991.
- Brandt, Søren (1993): *Knasegrynteranalyseprogrammer*, Weekendavisens bogtillæg, Bøger, Nr. 17-1993: 13.
- DANLEX 1987a: The DANLEX Group (Ebba Hjorth, Bodil Nistrup Madsen, Ole Norling-Christensen, Jane Rosenkilde Jacobsen, Hanne Ruus): *Descriptive Tools for Electronic Processing of Dictionary Data*. Studies in Computational Lexicography, Niemeyer, Tübingen 1987.
- DANLEX 1987b: DANLEXgruppen (Gert Engel, Ebba Hjorth, Bodil Nistrup Madsen, Ole Norling-Christensen, Jane Rosenkilde Jacobsen, Hanne Ruus; red.): *Ordbøger i Danmark. En oversigt*. København 1987.
- DANLEX 1989: DANLEXgruppen (Ebba Hjorth, Bodil Nistrup Madsen, Ole Norling-Christensen, Hanne Ruus, Jørgen Schack, Peter Widell): *Behov og Brugerprofiler for Dansk Ordbank*. København 1989.
- Dansk Ordbog 1989: Ebba Hjorth, Iver Kjær, Ole Norling-Christensen og Hanne Ruus: *Dansk Ordbog. Plan for en ordbog over dansk Nutidssprog*. 18. december 1989.
- Galberg Jacobsen, Henrik (1987): *Retskrivningsreglerne*, i Selskab for Nordisk Filologi, Årsberetning 1985-1986: 32-35. København 1987.
- Kristensen, Kjeld (1994): *Den Danske Ordbogs tekstkorpus og spORDhunde* (i denne bog).
- ODS: *Ordbog over det Danske Sprog*. Udgivet af Det Danske Sprog- og Litteraturselskab. Bind 1-28. København 1918-56.
- RO: *Retskrivningsordbogen*. Udgivet af Dansk Sprognævn. København 1986.
- V&B: Vinterberg, Hermann og C.A. Bodelsen: *Dansk-engelsk Ordbog*. 3. udg. ved Viggo Hjørnager Pedersen. København 1990.