

*Peter Widell og Mette Kunøe (udg.):
10. Møde om Udforskningen af Dansk Sprog
Århus 2004*

ODS, ODS-S OG FREMTIDEN

Af Else Bojsen og Lars Trap-Jensen (Det Danske Sprog- og Litteraturselskab)

1. ODS, Ordbog over det danske Sprog, "Den store danske ordbog", med supplement, ODS-S

ODS regnes for at være en af de store nationalordbøger i Vest- og Nordeuropa. Genremæssigt deler den status med bl.a. Grimms Deutsches Wörterbuch (DWB), Oxford English Dictionary (OED), Svenska Akademiens Ordbok (SAOB) og Het Woordenboek der Nederlandsche Taal (WNT). Men mens de andre går ned til eller over middelaldergrænsen, har ODS år 1700 som startpunkt. Den omfatter altså et kortere tidsrum end de andre. Og til forskel fra SAOB er det danske ordbogsværk nu færdigt. Femte og sidste bind af Supplementet ventes at udkomme i begyndelsen af 2005. Da 1. bind af Supplementet udkom i december 1992, skrev hovedredaktøren, Anne Duekilde, i en pressemeddelelse:

De 28 ODS-bind udkom 1918-56. De beskriver det danske rigssprog som det var dengang, og som det havde udviklet sig siden ca. 1700. Det der især fylder siderne i de mange bind er de udførlige eksempler på hvordan ordene er blevet brugt i tidens løb. I ODS er der ikke bare en forklaring til ordenes anvendelser, der er også illustrerende citater fra skønlitteratur og faglitteratur, fra aviser, love, bibeloversættelser og meget andet. I et sådant ordbogsværk står læseren over for helt konkrete eksempler på ordenes brug. Man møder kendte og ukendte tekster, lige fra Chr. V's Danske Lov over Holberg, H.C. Ørsted, H.C. Andersen, Søren Kierkegaard, St.St. Blicher osv. osv. til forfattere og opslagsværker fra vort eget århundrede.

ODS er altså en deskriptiv ordbog baseret på citater. Den beskriver sproget i en afgrænset historisk periode, men samtidig beskriver den jo samtidsproget, det sprog som var redaktørernes. Det sprog der var samtidsprog i 1918 da 1. bind af ODS udkom, var på mange punkter blevet historisk for ikke at sige forældet i 1954 da bind 27 udkom. Det er ikke blevet mindre forældet nu, 50 år efter, ved afslutningen af Supplementet. Supplementet dækker jo samme periode som hovedordbogen. Da den dominerende kildetype i ODS er skønlitteraturen, kan man sige at ordbogen er lige så forældet som litteraturen, men det er jo noget sludder! Så hellere citere den anden af Supplementets "gamle redaktører", Svend Eegholm-Pedersen: "ODS er ubestrideligt en smuk litteratur- og klassikerordbog med fyldig repræsentation af periodens hovedforfattere" (Eegholm-Pedersen 1993:71).

Men der er naturligvis visse områder der rammes hårdt af forældelse. Det gælder klædedragt og mode, livsstil, herunder også fx boligindretning, spisevaner, madretter, menneskesyn, det vi nu kalder politisk korrekthed, eller rettere manglen på politisk korrekthed. I ODS er en neger stadig en neger, og jazz kaldes "negermusik". Den 25. september i år var der et læserbrev i Politiken om hvor længe man har brugt ordet *bøsse* i betydningen 'homoseksuel mand'. Skribenten skriver: "Lis Jacobsens Ordbog Over Det Danske Sprog fra 1920'erne nævner ikke udtrykket, men det skyldes måske berøringsangst?" Havde læseren nu vidst lidt mere om ordbogen, ville han have kunnet se efter i Supplementet hvor den betydning af ordet er medtaget og med ældste belæg fra 1855. Berøringsangst er alt for nyt og moderne for ODS-perioden. Det samme gælder den opfattelse at spædbørn er lækre. Under supplementet til *Løjert* i betydningen 'svøb som man bandt spædbørn ind i' finder vi følgende citat fra 1935:

"smaa Barnepiger, der skævlæber sig paa savlende, vandladende, skaldede Rollinger, som i deres Løjertes ligner Ærkebiskopper i Dværgformat".

Måske begyndte synet på babyer at ændre sig i takt med at de ikke bare kom dumpende på de mest ubejlejlige tidspunkter, men kunne planlægges, ja ligefrem begynde som *reagensglasbørn*? Dette ord er belagt fra 1950 i betydningen 'børn undfanget ved insemination'.

Anden verdenskrig og besættelsestiden i Danmark har sat sit præg på periodens begreber og sprog. Det gælder selvfølgelig en masse udtryk der har direkte relation til den politiske situation. Men det gælder også indirekte det ordforråd der afspejler folks utryghed. For eksempel dukker der en masse nye sammensætninger op med ordet *hygge* som førsteled, fx *Hyggeaften*, *-belysning*, *-krog*, *-lampe*, *-musik*, *-pianist*, *-spreder*. Ved *Hyggebelysning* er der anbragt et citat fra Husmoderleksikon, 1948, som på sin egen lidt kliniske måde giver en definition på ordet:

Hyggebelysning, kunstig Belysning .. der ikke skal tjene som Arbejdsbelysning, men tilsigter at bidrage til en behagelig eller festlig Stemning i Rummet ved Hvile, Samtale, Selskabelighed el. lign.

Fra 1940 har vi et citat med *Hyggeonkel*:

Varmetante og Hyggeonkel .. i disse lange og kølige Aftener.

I mange tilfælde har vi været nødt til at genbruge et citat hvis der har været flere interessante ord i det. Men det er heldigvis sjældent at vi har måttet være så drastiske som i tilfældet *Varmetante* i bd. 5 af Supplementet. Der kommer til at stå: *Varmetante*, se u. *Hyggeonkel*.

Apropos *Varmetante* – det er til tider forbløffende hvad Supplementet bringer som nyheder. For eksempel er ordet *Radiator* først kommet med i Supplementet. Centralvarme er dog ikke et begreb der er opstået sent i ODS-perioden. Men man har været meget tilbageholdende på fremmedordsområdet, og det har så ramt radiatoren. Som opslagsord, vel at mærke, for i definitionen til en af betydningerne af ordet *Varmeapparat* har redaktionen brugt "radiator"

som synonym. Det har naturligvis også hændt at vi supplementsredaktører¹ har måttet benytte et forklaringsord som ikke har været kendt i ODS-perioden. Ved ordet *Vaskesuger* i Supplementet synes vi det var naturligt at give det moderne synonym "svupper". Vi har dog afholdt os fra også at nævne det mere slangprægede *kamelkys*.

Man kunne også have lyst til at nævne hvordan ordet *Banan* er suppleret. Flere af redaktørerne hører til de årgange som spændt ventede på at stifte bekendtskab med denne stærkt opreklamerede frugt efter krigen (og blev grumme skuffet). I ODS fylder artiklen kun 8½ linje, men i Supplementet 104 linjer (inkl. sammensætninger), jf. illustrationen i Svend Eegholm-Pedersens føromtalt artikel (Eegholm-Pedersen 1993:77). Det er dog ikke alle ord der har fået så kraftigt et supplement.

Et lille glimt fra modens verden bør også gives her. Af supplementets nytilkomne ord kan nævnes *Roll-on*. Nej, det er ikke en deodorant, og det har heller ikke noget at gøre med et fragtskib. Det er et stykke dameundertøj. Mormor gik med korset hvorpå der sad en række strømpebånd til at holde strømperne oppe med. Men mor var moderne – hun havde roll-on: et elastisk rør der strammede maven ind og i øvrigt var forsynet med strømpebånd. Datteren gik med livstykke hvor strømpebåndene var knuppet på, og som teenager med strømpeholder, som nærmest var et bælte med strømpebåndene hængende ned fra. Indtil revolutionen kom – i form af *Strømpebukser*. Dette ord er også kommet med i Supplementet, med en lidt speciel forklaring. Der vil komme til at stå:

Strømpebukser, pl. (isoleret ældre belæg, se ndf. – nyere brug (jf. eng. (amer.) panty hose) fra slutningen af 1950'erne) bukser og strømpeud i ét.

Ældste citat er fra 1842, hvor det drejer sig om en slags oversættelse af det franske ord *pantalon*, dvs. lange bukser der sidder lige så tæt til benet som strømper. Og ellers henvises der til Pia Jarvads Nye Ord 1955-98, med forklaringen "beklædning for kvinder og børn".

Og alligevel er der jo intet nyt under solen. I ODS finder vi betegnelsen "forældet" ved et helt tilsvarende ord mht. orddannelsen: *Sokkebrog*, som forklares som "gl. nordisk mandsdragt, bestaaende af benklæder, der nåede helt ned om fødderne og saaledes tillige tjente som sokker". Citatet lyder: "Gunløg .. var iført .. hvide Sokkebrog." Det stammer fra N.M. Petersens oversættelse af Gunløg Ormstunges Saga og er således litterært.

Om ODS og ODS-S kan man sige med endnu et citat fra Anne Duekildes pressemeddelelse fra 1992:

Et sprog får de ord som der er et behov for. Sproghistorien afspejler kulturhistorien, og en ordbog med gode citater er spændende læsning for enhver der er interesseret i sprogets

¹ dvs. til bd. 4 og 5 foruden de nævnte "gamle redaktører" Henrik Andersson, Else Bojsen og Vibeke Winge.

eller kulturens historie. En sådan ordbog er naturligvis uundværlig for den der vil forstå ordene i gamle tekster eller har brug for at vide hvordan man gennem tiden har kunnet udtrykke sig i dansk.

2. ODS og fremtiden.

Nu hvor sidste punktum så er sat for femte og sidste supplementsbind, er det måske lidt vemodigt at fastslå at en epoke er slut. Helt slut? Nej, til trost for de mange trofaste fans (ordet *Fan* optræder i ODS-Supplementet med første belæg fra 1928) er vi glade for at kunne oplyse at der også bliver en plads til ODS i den digitale tidsalder. Det er lykkedes Det Danske Sprog- og Litteraturselskab (DSL) at skaffe midler til et elektronisk projekt der bl.a. omfatter digitalisering af ODS og publicering af en elektronisk, samlet version af ODS og Supplementet. I første række stiler vi mod en online-udgave på internettet, men også en cd-rom-udgave kan blive en mulighed.

På høje tid, vil mange nok sige, og det er da også en sag der har ligget DSL på sinde gennem lang tid. Nu hvor det endelig er så vidt, må vi vende det til noget positivt og i stedet glæde os over at vi i hvert fald kan lære af de dyrekøbte erfaringer som vores søsterprojekter og pionerer på området har høstet. De fire tidligere nævnte ordbøger er nemlig alle digitaliserede i dag, og nogle af dem var endda meget tidligt ude: SAOB begyndte i 1983, OED og WNT i 1984, mens DWB som det sidste projekt blev digitaliseret i perioden 1998-2003/4.

Den første beslutning der har skullet træffes, har været digitaliseringsmåden: scanning eller indtastning? Her er virkelig noget at lære fra de andre ordbøger: For SAOB, der som den eneste valgte scanneløsningen, strakte digitaliseringsperioden sig over 15 år, og alene tidsrammen har gjort at vi ikke har været i tvivl om at *indtastning* er en samlet set både billigere og mere pålidelig metode. Vi har derfor valgt at følge en model som vi vil skitsere i det følgende.

2.1 Indtastning og projektforsløb

Modellen lægger sig tæt op ad den der har været brugt af det seneste af de nævnte projekter, DWB: dobbeltindtastning af manuskriptet. Vi samarbejder om projektets første fase med Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften, der er en afdeling ved universitetet i Trier, og dem der var ansvarlige for digitaliseringen af DWB. Selve processen kan inddrages i to faser: en første fase hvor rådigitaliseringen finder sted, og en efterfølgende fase hvor det digitale manuskript fortolkes og opmærkes strukturelt. Skematisk kan det sammenfattes i følgende figur (der er en modificeret version af den tyske projektskitse).

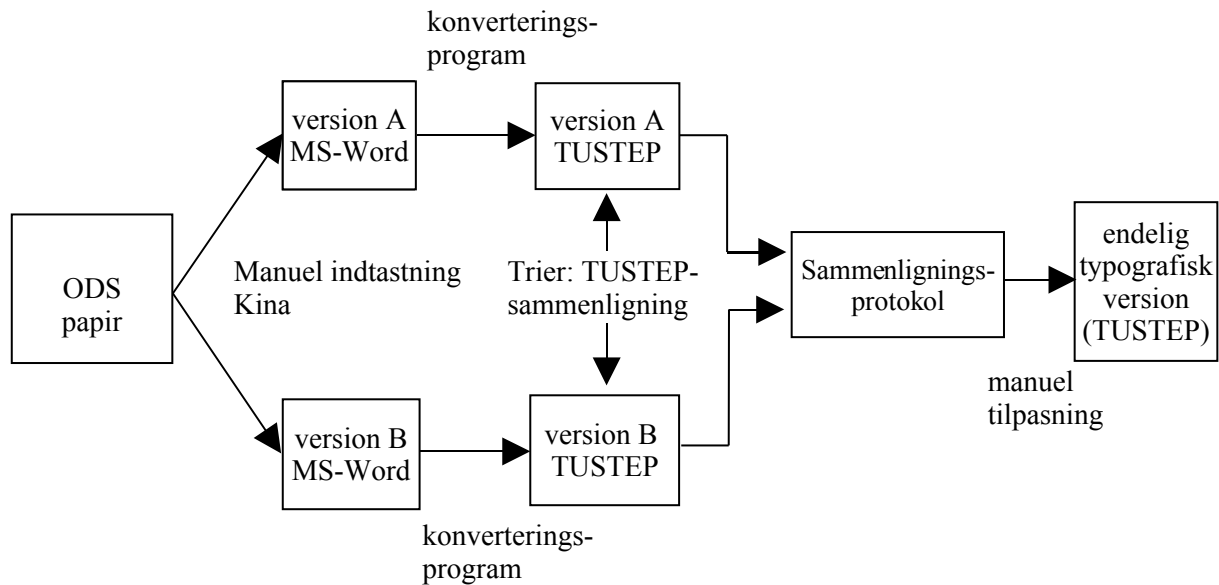


Fig. 1 Rådigitalisering (Kina, Trier)

Forud for selve indtastningsarbejdet har vi analyseret ODS i detaljer og udarbejdet en indtastningsmanual for at sikre at alle typografiske oplysninger fra den trykte udgave bevares. Manualen indeholder anvisninger på hvilke koder indtasterne skal skrive for de forskellige specialtegn, symboler og andet der ikke er indeholdt i standardversioner af tekstprogram og tastatur. Den fysiske indtastning bliver udført af firmaet TQY DoubleKey i Nanjing, Kina, efter *double keying*-metoden, hvor teksten indtastes i to versioner af to uafhængige grupper indtastere i et almindeligt tekstbehandlingsprogram (in casu: MS Word). Efterfølgende konverteres de to versioner til et såkaldt TUSTEP-format², og de to versioner sammenlignes automatisk. Forskellen mellem de to versioner udskrives i en sammenligningsprotokol, og denne protokol bearbejdes endelig manuelt indtil man har den endelige typografiske version, en version der har form som tekstfiler hvor tekstens formatering er opmærket med de særlige, maskinlæsbare TUSTEP-koder. Fig. 2 giver et indtryk af hvordan filerne ser ud.

```

$0303.01 <P>__<A+1>#F+afstr#.omme,#F-</A+1> #/+v.#/- [#P+!#P-au#P+^!#P-
sdr%::om'#P+^e#P-] #/+#s+str#.omme
$0303.02 bort.#/-#s- #/+vist kun som vbs.:#/- <A+1>{shu}
#F+Afstr#.omning,#F-</A+1>
$0303.03 en. Nedb#.orens . . Afstr#.omning gennem
$0303.04 #/+(dr#.^onledninger). LandbO.I.608.#/- #F+#.#F- #/+(konkr.)#/-
$0303.05 Afstr#.omningen gennem Dr#.^anledningerne
$0303.06 #/+(udgjorde)#/- ca. 31 p. Ct. af Nedb#.oren. #/+smst.
$0303.07 jf. (overf.; sj.):#/- her #/+(#P+O#P-: i London)#/- er Moral,
$0303.08 der kan ikke dv#.^ales ved enkelte Udv#.^axter
$0303.09 og Afstr#.omninger, som altid findes i en
$0303.10 stor By. #/+HCAnd.XII.193.#/-</P>
  
```

²TUSTEP (= Tübinger System von Textverarbeitungsprogrammen) er et særligt tekstbehandlings- og satssystem til filologiske publikationer

```

$0303.11 <P>__<A+1>#F+afstudse,#F-</A+1> #/+v.#/- [#P+!#P-au#P+^!#P-
sdus#P+^e#P-] #F+-ede.#F- #/+vbs.#/- #F+-ning.#F-
$0303.12 #/+(1. br.) (fjerne ell.) #s+afkorte#/-#s- #/+ved #s+studsning.#/-
#s-
$0303.13 Afstudse . . Tr#.^aer. #/+vAph.(1759).#/- En Hund
$0303.14 med afstudsede #.Oren. #/+VSO.#/- Dykkerser
$0303.15 (S#.om med afstudsede Hoveder). #/+Hallager.
$0303.16 231.#/- kort Haar og afstudset Sk#.^ag. #/+Ing.
$0303.17 DM.196. H#.oyen.Moltke.4.#/-</P>

```

Fig. 2 Artiklerne *afstrømme* og *afstudse* i TUSTEP-format

Fordelen ved denne fremgangsmåde er dels at alle de typografiske oplysninger fra det trykte forlæg bevares – og det er meget vigtigt for strukturfortolkningen – dels at dobbeltindtastningen sikrer at fejlprocenten bliver så lav at en efterfølgende manuel korrekturlæsning kan undværes. Det sidste er i sagens natur altafgørende for tids- og resurseforbruget for et værk af ODS' omfang. Stikprøver fra DWB har vist at indtastningen har ført til at det elektroniske manuskript stemmer 99,997% overens med det trykte forlæg, og eftersom ODS med hensyn til såvel kompleksitet som typografisk tydelighed er noget klarere end DWB, har man stillet os i udsigt at vi kan nå op på 99,999% (dvs. 1 fejl pr. 100.000 typeenheder).

Når den endelige typografiske version er modtaget fra Kina og Tyskland, begynder DSL's del af arbejdet, nemlig strukturfortolkningen af den typografiske opmærkning. Den kan i skematisk form anskues på følgende måde:

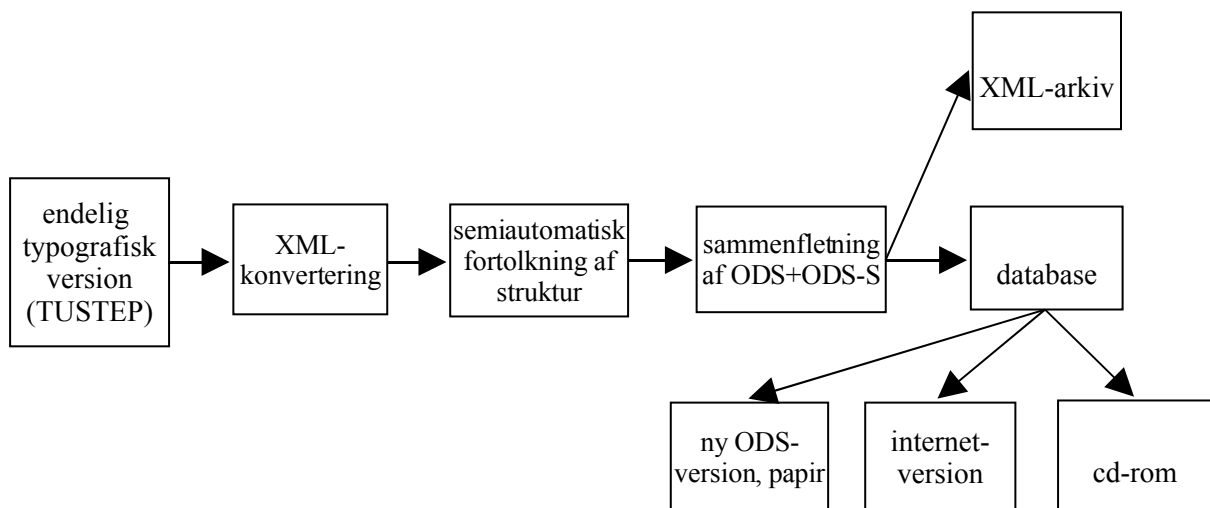


Fig. 3 Fra rådata til fulddigitaliseret version

Det siger næsten sig selv at den typografiske information er helt afgørende når man skal bevæge sig fra typografisk opmærkning til strukturopmærkning. Det er ganske enkelt den eneste måde man kan foretage en indholdsmæssig fortolkning af artiklerne på. Skriftsnit, skriftgrad, linjeskift, indrykning, ordinær, kursiv, fed el. spatieret skrift er, ofte i kombination med inter-

punktionstegn, de signaler bogversionen råder over til at angive hvornår én oplysningstype begynder og en anden slutter. Endnu er projektet ikke så vidt fremskredet at vi ved hvilke udfordringer der venter os på dette punkt, ligesom det bliver interessant at finde ud af hvor langt ned i indholdsstrukturen det er muligt at komme med automatiserede operationer. Så snart det bliver nødvendigt at en redaktør gennemgår manuskriptet manuelt og retter det til, bliver det hurtigt meget resursekrævende. Den strukturelle fortolkning bliver den ene hovedopgave i de kommende år.

Den anden hovedopgave består i at flette det elektroniske manuskript sammen med ODS-Supplementet. Supplementet foreligger jo allerede i elektronisk form og har derfor ikke skullet omkring Kina og Tyskland først. Af samme grund kan Supplementet også være en hjælp til den indholdsmæssige fortolkning af typografien, men det kan ikke give det endelige svar. For det første er nemlig Supplementet i nogen grad blevet ensrettet i sin struktur til det elektroniske format så strukturen er mere homogen end man kan se det i ODS. For det andet er strukturen i Supplementet primært fastlagt med henblik på papirpublicering; den er ikke i et format der overholder en standardiseret dokumentbeskrivelse, fx i form af en DTD.

Det vil derfor være naturligt at begge værker i forbindelse med strukturopmærkningen konverteres til tidens standard for dokumentbeskrivelse, XML. Herefter kan de to manuskripter flettes sammen i en fælles base, dels i et XML-arkiv der sikrer langtidslagring af selve dokumentindholdet, dels i en relationel database som kan danne udgangspunkt for såvel avancerede søgninger i en online-udgave som for andre former for publicering – hvad enten det er elektronisk eller på papir.

Opgaven med at flette de to værker sammen er en udfordring der ikke bør undervurdes. Så længe det drejer sig om at tilføje hele artikler, kan det forholdsvis nemt gøres automatisk, men supplementsartiklerne dækker jo over en bred vifte af tilføjelser, ændringer og sletninger og kan gælde et hvilket som helst element i en eksisterende artikel. Fra vores kolleger på OED ved vi at det var en af de helt store udfordringer ved digitaliseringen af OED, og nogle af problemerne har forfulgt dem helt indtil i dag. Og også her gælder det at de operationer der ikke kan udføres maskinelt, er ekstremt resursekrævende. Vores engelske kolleger oplyser således at digitaliseringen af OED i 1984-89 kostede 13,5 mio. dollars. Vi råder over væsentlig færre penge – selv målt i danske kroner – til vores projekt.

2.2 Tidsramme

Projektet er igangsat fra 1. januar 2004 og vil strække sig over en periode på 5½ år. Det vil sige at arbejdet skal være afsluttet og ODS være fuldt tilgængelig i en internetversion senest 1. juli 2009. Vi går dog ud fra at kunne præsentere offentligheden for noget væsentlig før. En af

fordelene ved internetpublicering er at det ikke behøver være det fuldt færdige produkt der lanceres med det samme. Man kan for eksempel tænke sig at vi allerede inden for to-tre år lægger en version ud som har begrænsede søgemuligheder, og hvor sammenfletningen med ODS-S endnu ikke har fundet sted. Brugeren vil med en sådan version få adgang til præcis de samme opslagsord som findes i papirudgaven, og oven i købet få både ODS- og ODS-S-artiklen på én gang som resultatet af en enkelt søgning, blot som to sideordnede søgeresultater i stedet for ét sammenflettet. Siden hen vil der kunne ske løbende revisioner i takt med at flere af deloperationerne blev færdiggjort.

Eftersom projektet stadig befinder sig i sin indledende fase, må ovenstående tages med det fornødne forbehold. Først når vi for alvor får taget fat på konverteringen af de typografisk opmærkede filer til XML og den tilsvarende konvertering af ODS-S-filerne, kan vi få et pålideligt indtryk af tidsforbruget og dermed have et bedre grundlag for at udtale os om en realistisk publiceringstakt.

Litteratur

Das Deutsche Wörterbuch von Jacob Grimm und Wilhelm Grimm, internetversion, www.dwb.uni-trier.de.

Eegholm-Pedersen, Svend (1993) ODS-Supplementet. *Nordiske studier i leksikografi*, nr. 2:69-77.

Het Woordenboek der Nederlandsche Taal, elektronisk version, se www.inl.nl/WNT.

Jarvad, Pia (1999) *Nye Ord. Ordbog over nye ord i dansk 1955-1998*. København: Gyldendalske Boghandel, Nordisk Forlag A/S.

Oxford English Dictionary, internetversion, www.oed.com el. <http://dictionary.oed.com>.

Svenska Akademiens Ordbok, internetversion, <http://g3.spraakdata.gu.se/saob>.