

Ømålsordbogen som leksikografisk, sproglig og kulturhistorisk resurse

Asgerd Gudiksen & Henrik Hovmark

As part of an extensive process of digitalization the previously edited volumes of *Ømålsordbogen* (The Dictionary of Danish Insular Dialects) (1–11, *a-lindorm*) are now available as one searchable xml file. We show how this new digital tool, even in its primitive state in terms of tagging, constitutes a valuable resource, which not only enables to explore the rich linguistic and cultural data in the dictionary in new ways but also saves time and brings more rigour to the editing process.

1. Ømålsordbogen anno 2017: omstilling og digitalisering

Ømålsordbogen (ØMO) er en af de store nationale videnskabelige dokumentationsordbøger i Norden (jf. fx *Norsk Ordbok*, *Ordbok över Finlands svenska folkmål*, mfl.). ØMO beskriver de traditionelle danske dialekter, dvs. talesproget blandt bønder og fiskere, på Sjælland, Lolland-Falster, Fyn og omliggende øer i perioden 1750–1950, med hovedvægten på tiden 1850–1920. Ordbogen medtager hele ordforrådet hos landbefolkningen og giver detaljerede beskrivelser af ordenes udtale, bøjning, syntaks, betydning og brug. Et særkende ved ordbogen er at den også giver indgående etnologisk-encyklopædiske beskrivelser af de redskaber, arbejdsprocesser og traditioner der hørte til ældre tiders bonde-, håndværker- og fiskerkultur, med henvisning til den folkelige terminologi der er knyttet til dem (jf. Gudiksen & Hovmark 2009). Under artiklen *bage* er der fx et afsnit på tre spalter om bagning i den gammeldags murede bageovn, og under *fastelavn* en skildring af dyster, optog, lege og gilder. Ordbogens oplysninger er sammen med de tilgrundliggende samlinger en vigtig del af dansk kulturarv.

Første bind af ordbogen udkom i 1992 efter en slags førstegenerationsdigitalisering i slutningen af 1980'erne, hvor man tog et særligt tekstbehandlingsprogram og et korpusværktøj i brug. Herefter er ordbogen udkommet i regelmæssigt tempo med et bind hvert andet år, senest bind 11 (*kurv-lindorm*) i 2013. Halvdel af materialet er således redigeret. I 2013 var en teknisk omstilling imidlertid påkrævet, og man gik derfor i gang med en ny digitaliseringsproces finansieret af Carlsbergfondet. Processen har omfattet flere forskellige dele: udarbejdelse af en dania-font i unicode, et nyt korpusværktøj (i samarbejde med Det Danske Sprog- og Litteraturselskab), digitalisering af de trykte bind 1–11 (*a-lindorm*) til xml-format, indkøb og opsætning af et nyt databasebaseret redigeringsværktøj (iLEX) inkl. udarbejdelse af xml-skema samt publiceringsløsning. Både database

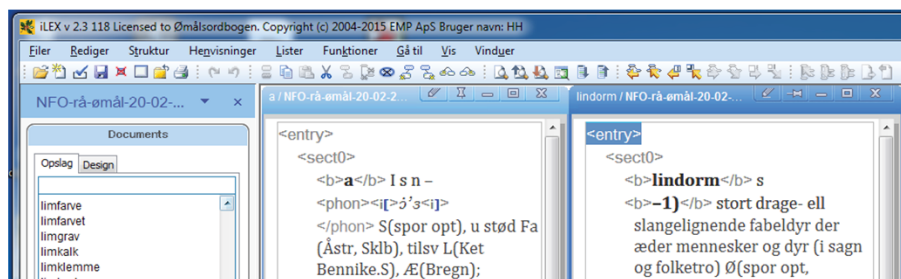
og publiceringsløsning forsøges så vidt muligt tilrettelagt så ordbogen fremover ikke er bundet til at udkomme på tryk, men også er forberedt til publicering på andre platforme. Hertil kommer at man er i gang med en digitalisering af ordbogens seddelsamling. Målet er en fulddigital udgivelsesproces. Bind 12 forventes at udkomme i 2018. Mange af de nye resurser foreligger indtil videre dog kun i forholdsvis rå former. Det gælder fx de digitaliserede trykte bind. De udgør dog allerede en værdifuld resurse, og i denne artikel vil vi give nogle eksempler på de muligheder som denne nye resurse giver.

2. Ømålsordbogen bind 1–11 som samlet, søgbar base

Da ØMO begyndte at udkomme i 1992, var det takket være udnyttelsen af det DOS-baserede tekstbehandlingsprogram T3, der rummede mulighed for at designe egne fonttegn. Det gjorde det muligt selv at udarbejde fonttegn i lydskriften *ania* og fremstille et manuskript med lydskrift, noget de daværende programmer ellers ikke gav mulighed for

T3 viste sig imidlertid også at have en bagside. Programmet kunne i praksis kun håndtere en lille datamængde ad gangen, typisk hvad der svarer til fx 10–15 sider i den trykte bog. Søgning i det samlede, udgivne materiale har derfor aldrig været muligt. Et lige så stort, og med tiden langt mere alvorligt problem var at filformatet havde sin egen, partikulære standard og kodning. De redigerede data blev hurtigt umulige at læse i og af andre programmer og kunne ikke umiddelbart konverteres. De redigerede data var dermed reelt ikke tilgængelige, og var heller ikke sikret i digital form.

Dette har heldigvis ændret sig. I forbindelse med omstillingsprocessen lykkedes det en medarbejder hos konsulentfirmaet EMP at knække koden til T3 ved hjælp af printfilerne, og data fra alle de trykte bind (*a–lindorm*), er blevet konverteret til unicode-tegn og foreligger nu som en let opmærket xml-fil som p.t. kan tilgås fra det nye redigeringsprogram, iLEX (ØMO-D). På Figur 1 ses således opslag på den første og sidste artikel i ØMO-D, *a* og *lindorm*, samt til venstre et udsnit af den alfabetiske ord- eller artikelliste.



Figur 1: Opslag i ØMO-D på artiklerne *a* og *lindorm* (første og sidste artikel i de hidtil trykte bind af ØMO).

En simpel søgning i basen kan foregå via ordhjulet til venstre i iLEX, men iLEX tilbyder langt flere søgemuligheder via et søgepanel i en funktionsmenu hvor både formelle søgeudtryk og ikke mindst xml-opmærkningen kan udnyttes til både søgning og statistik. Da basen med ØMO-D kun er minimalt opmærket efter skriftsnit ((fed), <i> (kursiv), <phon> (lydskrift – dania), <spatieret> (i forb s p a t i e r e t s k r i f t) samt i tre meget grove sektionsinddelinger (<sect0>, <sect1> og <sect2>) (jf. Figur 1), er søgning i fritekst ofte den mest oplagte søgestrategi, men undertiden kan opmærkningen dog også være til hjælp til afgrænsede eller kombinerede søgninger (jf. afsnit 3).

3. ØMO-D som leksikografisk resurse

Flerordsforbindelser består som navnet siger af mere end ét ord, og en bruger kan komme i tvivl om hvor en sådan forbindelse behandles i en ordbog. Løsningen er normalt at en forbindelse kun behandles ét sted, dvs. i den traditionelle ordbog ordnet.dk efter enkeltord: i én artikel, men ofte med henvisning fra de andre artikler hvor forbindelsen potentielt kunne have været behandlet (idet funktionsord o.l. normalt ikke tæller med som potentielle opslagsord, fx Svensén 2004:247ff.). Dette princip følges også i ØMO, hvor princippet har en ekstra begrundelse, nemlig at der skal gives en samlet udbredelse for en flerordsforbindelse dér hvor den behandles, dvs. fx:

lus s ... i forb k n æ k k e l u s Ø(spor opt)
knække v ... i forb k n æ k k e →l u s

For at kunne overholde princippet om at beskrivelse og udbredelsesangivelse kun gives ét sted, skal man have alle relevante belæg til rådighed når man går i gang, også hvis den artikel man står i, skal resultere i en henvisning og ikke en fuld beskrivelse (fx hvis man er i gang med at redigere artiklen *knække*). Behandlingen af flerordsforbindelser i ØMO stiller imidlertid redaktørerne over for nogle særlige udfordringer på grund af den måde som datamaterialet er organiseret på. Datamaterialet er indtil videre organiseret i en række forskellige delsamlinger, og en forbindelse kan være behandlet i flere samlinger. Tidligere redaktører var dog naturligvis opmærksomme på denne problemstilling, og derfor er mange belæg allerede forsøgt samlet under ét opslagsord – fx er ordsprog og talemåder, som ofte kan have en kompleks struktur med mange potentielle opslagsord og mange varianter, samlet i én særlig delsamling. Alligevel kan der gemme sig materiale flere steder. Der er også forskel på i hvor høj grad den enkelte redaktør har opfattet en flerordsforbindelse som fast og dermed behandlet den eksplicit som en forbindelse, eller fx ”blot” har medtaget forbindelsen som et citat. Og ulykkeligvis har man heller ikke haft knivskarpe principper for under hvilket ord (fx ”første substantiv”, ”første substantiv eller verbum” e.l.) en forbindelse skulle anbringes

i hverken samlinger eller ordbog: Man fulgte til en vis grad ODS, men ODS's organisering af forbindelser følger netop ikke noget entydigt princip af den netop nævnte type. Det kan altså være svært at vide om en flerordsforbindelse evt. allerede er blevet behandlet i de tidligere trykte bind.

Omstillingsprocessen vil kunne råde bod på denne vanskelighed, som i den nuværende produktionsproces ofte koster mange kræfter og tid for redaktørerne i form af gentagne kontrolopslag i seddelsamlingen og de trykte bind. For det første er der håb om at man fremover vil kunne søge alfabetisk på tværs af alle delsamlinger når digitaliseringen af seddelsamlingerne er afsluttet. For det andet giver ØMO-D allerede nu mulighed for at lave samlede undersøgelser af hvor og hvordan ord og ordforbindelser evt. er blevet behandlet i de tidligere trykte bind.

Ordet *lus* er forholdsvis velbelagt i ØMOs samlinger, og det seddelmateriale som havde *lus* som opslagsord, og som derfor dannede udgangspunkt for redaktørens arbejde, viste også at ordet indgår i en del ordsprog og talemåder. ØMO-D blev derfor taget i brug for at tjekke evt. yderligere relevant materiale i de trykte bind. Undersøgelsen viste hurtigt at der gemte sig materiale, ikke kun som udhævede forbindelser, men også i fx citatstof. I materialet under *lus* lå fx et belæg på talemåden *være rig på lus (og lopper) og lange negle*. Denne flerordsforbindelse rummer flere ord som potentielt kunne fungere som opslagsord, men kun ét, nemlig *lang*, lå tidligere i alfabetet. Et simpelt opslag på ”lange negle” viste imidlertid at talemåden ikke var behandlet under *lang*, men til gengæld at en variant var medtaget under *fattig*, som der nu bliver henvist til under *lus*. Søgningen fandt i øvrigt også en helt anden talemåde i artiklen *hip* (II), hvor *lus* var kombineret med *lange negle* – som ikke lå i materialet til *lus*.

Problemet ligger dog ikke kun i at der er vanskeligheder forbundet med at finde alt relevant stof og koordinere det konsekvent. Selve stoffet kan undertiden også være præget af meget stor variation. Fx var der i materialet til *lus* belæg på ordsproget *der er liv i kludene, sagde manden, han pillede lus af sin frakke*. Ordsproget havde varianten *hun skal nok få sat liv i kludene, som kællingen sagde, hun pillede lus af uldklokken*. Et opslag i ØMO-D viste at ordsproget – eller ordspørgene? – allerede var behandlet under *klud*, dog med mangelfuld udbredelse. Under *klud* fandtes imidlertid en tredje variant: *der er liv i kludene, sagde pjaltekræmmeren, lusene krøb omkring i pjalterne*. I denne slags tilfælde kan det være svært at pege på én grundlæggende form af et ordsprog eller en talemåde (som ville kunne styre udvælgelsen af et hovedopslagsord efter mere objektive kriterier).

Den meget specifikke søgning på ”lange negle” gav to meget konkrete resultater. En søgning på formen ”lus” og/eller på bøjningsformer heraf gav betydeligt flere resultater (over 200). Den sparsomme opmærkning efter skriftsnit i ØMO-D giver dog mulighed for at begrænse omfanget af irrelevant støj. En søgning inden for tagget <i> vil således begrænse søgningen til tekst i kursiv, hvilket fx vil omfatte citater og henvisninger i ØMO, og en søgning der inddrager tagget (’fed’) vil give opslagsord og orddannelser som søgeresultat. Selvom den rene tekstsøgning på formen ”lus” med evt. bøjningsformer gav forholdsvis mange

resultater, viste det sig alligevel at være umagen værd at gennemgå resultaterne som altså kunne forekomme alle steder i en artikel. Ordet *lus* optrådte fx i en af definitionerne til ordet *bid* (I.8). Dette ord forekom ikke i materialet til *lus*, men brugtes metonymisk og uformelt om *lus*, og det vil være relevant at henvise til ordet og dermed koble materialet sammen i basen.

Samlet set bidrager ØMO-D med værdifulde oplysninger som sætter redaktørerne i stand til at gøre de fremtidige artikler mere konsekvente og udtømmende i forhold til det komplekse materiale. Muligheden for at søge elektronisk i ØMO-D er tidsbesparende og giver et mere fuldstændigt datagrundlag at arbejde ud fra.

4. ØMO-D som sproglig resurse

ØMO har som sagt i sin egenskab af en stor dokumentationsordbog detaljerede beskrivelser af udtale, bøjning, syntaks, betydning og brug. Kildegrundlaget afviger på mange måder fra den typiske dokumentationsordbog, en ordbog over standardsproget som fx *Ordbog over det danske Sprog* (ODS). Ordbøger over standardsproget som ODS beskriver overvejende skriftsproget og bygger på excerpter af trykte kilder og – i nyere tid – på elektroniske korpora. Dialektordbøger som ØMO inddrager også trykte kilder, fx bondedagbøger, dialektlitteratur, etnologiske og folkloristiske skildringer. Men de vigtigste kilder til ordbogens seddelsamling er ordbogens eget indsamlede materiale. Materialet hidrører fra direkte optegnelser med dialekttalende meddelere foretaget af dialektforskere, meddelernes besvarelser af spørgelister og endelig ordsamlinger og ordbøger udarbejdet af lokale dialekttalende. Der er tale om et meget heterogent materiale der må vurderes kildekritisk i forhold til den situation det er blevet til i. Optegneren kan fx misforstå noget, eller meddelelserne kan huske forkert (ofte drejer det sig om forhold der ligger mange år tilbage i forhold til optegnelsestidspunktet). Det kræver altså en omfattende baggrundsviden at forstå oplysningerne fuldt ud. Mange sprogforskere vil derfor være mere tjent med de generaliseringer i ordbogen som redaktørerne efter kritisk vurdering er nået frem til end det måske lidt fortegnede billede de selv ville være nået frem til ved at gå direkte til kilderne.

Det har fra begyndelsen da også været tanken at ordbogen skulle kunne danne grundlag for (mindre) sproglige undersøgelser (jf. Pedersen 1994). Indtil for nylig måtte man gennembladre ordbogen for at samle materiale til en sproglig undersøgelse, som fx en undersøgelse af variation i verbalbøjningsmønstre (Gudiksen 2011) – en ret arbejdskrævende opgave, men dog overkommelig i modsætning til at skulle have etableret det samme overblik ved hjælp af grundmaterialet. Men vores nye resurse giver nye muligheder for sproglige undersøgelser. Ved hjælp af basen kan man fx lave visse syntaktiske undersøgelser. Ganske vist kan man ikke søge systematisk på syntaktiske konstruktioner, men man kan søge på grammatiske markører som ”upers” (dvs. upersonlig), ”i sb anv” (dvs. i substantivisk anvendelse), ”i adj anv” (dvs. i adjektivisk anvendelse) eller ”sideordnet”. En

hurtig søgning på ”upers” bringer bl.a. en anvendelse af *kaste* som er ukendt i standardsproget, nemlig ’blæse med kraftige vindstød’, der bl.a. er afhjemlet på Tåsinge med citatet *det er slemt til at kaste fra nordvest* (*kaste* III.6.1). Man kan på tilsvarende måde søge på betegnelser for lydlige fænomener som ”kortvokalstød”, ”v-stød”, ”enklitisk”. En søgning på ”enklitisk” fører bl.a. hen til konstruktionen *til den* ’afsted’, fx *vi skal te ’en* ’vi skal afsted’ (under *den* 2). Et andet søgeresultat er citatet: *Nu har jeg knaldfyret en hel klokke time og endnu er hæ’d varmt*, en sammentrækning af ordet *her* med nægtelsen *itte* (en sideform til *inte*). Desuden fremkommer også den generelle beskrivelse af sådanne enklitiske forbindelser med nægtelsen under udtaleangivelserne ved *itte* (II), fx [*sgadə*] ’skal itte’.

5. ØMO-D som kulturhistorisk resurse

Omkring 1940 blev ØMOs seddelsamling opdelt i de tidligere omtalte delsamlinger, der – foruden delsamlinger som ordsprog og talemåder – omfatter emnesamlinger som fx *høst*, *pløjning*, *bagning* (og en restsamling for den del af ordforrådet der ikke kunne indpasses i emnerne). Formålet var at kunne foretage en samlet analyse af ordforrådet inden for de enkelte begrebsområder til brug for de etnologiske beskrivelser, der som tidligere nævnt er en central del af ordbogen, og udarbejde synonymkredse, herunder udvælge et hovedsynonym hvor alle synonymerne er samlet, og hvortil der henvises. *Lokum* er et eksempel på et hovedsynonym inden for emnet ”gården og dens indretning”. Artiklen afsluttes med synonymkredsen *das*, *hus*, *kloset*, *retirade*, *skidehus*, *vandhus*, *wc*. Der er altså også synonymer længere fremme i alfabetet end redaktionen er nået til, takket være forredaktionen af emnerne, og en søgning på *lokum* har ikke givet yderligere eksempler inden for det redigerede. Men artiklen *lokum* har inspireret os til en anden søgning der har bragt viden frem der lå og gemte sig i artiklerne.

Betydningsbeskrivelsen i artiklen *lokum* slutter med følgende måske lidt overraskende oplysning: ”indført fra omkr 1860, i ældre tid brugte man normalt grebningen i stalden”, og det inspirerede til at søge på ”indført”. Resultatet var 26 artikler hvor ”indført” indgår et eller andet sted i artiklen. Nogle af eksemplerne havde ikke relevans i denne sammenhæng, fx hvor ”indført” indgår i et citat. Men tilbage blev der 12 artikler med redaktionel markering. I tre tilfælde er der angivet et præcist årstal, nemlig hvor noget er indført ved lov: *femogtyveøre* (1873), *jagt-tegn* (1922) og *konfirmation* (1736.) I de øvrige tilfælde viser angivelserne at der er tale om gradvise ændringer. Det er karakteristisk for datidens bondesamfund at nye ting vinder frem i langsom takt. Det angives fx at *højstade* (en type bihus af træ) er indført ”ca. 1860”, *kødhakkemaskine* ”omkr 1880–1900” og *kalvesteg* ”i takt med indførslen af *komfuret*”.

Søgning på ”først alm”, ”først i brug”, ”først brugt” gav yderligere eksempler på forandring af bondekulturen, bl.a. *henkoge*, *høsegård*, *hårbørste*, *hårnål*, *jordbær*, *juletræ*. I mange tilfælde drejer det sig dog om ting som inden for andre

sociale lag har været kendt tidligere, men altså først bliver en del af bondekulturen i slutningen af 1800-tallet. ”Nyere” var topscoreren med 625 artikler og ”yngre” med 593 artikler. ”Nyere” bruges undertiden om udtalemæssige forhold, og ”yngre” ofte.

Under ”nyere” er der ikke kun ord for ”nye” sagforhold som *lokum*, *henkoge* og *jordbær*, men også nye ord for gamle ”sager”, fx: *kostald*, der afløser fx *nøds*, *kohus*, *kolade* (idet *stald* oprindeligt kun blev brugt om hestenes afdeling), og *idiot* brugt som skældsord (oprindeligt blev *idiot* kun brugt som betegnelse for en svært mentalt retarderet person). Arbejdet med at nå frem til tidsfæstelse af ord og fænomener er noget vi bruger en del tid på, og det er derfor glædeligt at man med de nye søgemuligheder hurtigt kan samle sådanne oplysninger sammen og nå frem til nye indsigter som vel at mærke ikke fremgår af de tidligere emneredaktioner. Søgning på ”nyere” og ”yngre” vil givetvis også kunne give muligheder for at etablere et mere nuanceret billede af den rigsmålspåvirkning af udtalen der sætter ind inden for ordbogens periode.

Citaterne i ØMO er ofte ikke kun sprogligt, men også kulturhistorisk interessante. Undertiden giver de også et indblik i holdninger og mentalitet i det gamle bondesamfund, noget vi naturligvis også lægger vægt på ved udvælgelsen af citater. I materialet til opslagsordet *københavnsk* var der to citater som blev udvalgt af den grund: *hun er bleven så kevihavnsk at hun er færdig at brække tungen* og *hun er blevet så københavnsk i de par måneder så hun næsten ikke er til at forstå*. Men kan billedet suppleres ved søgning på *København*? Ja, under *kloges* finder man fx citatet: *je ku jonok kloves ae* (kloges af, dvs. ’forstå’), *a ded vaa saan in rejti Københavner Sjaafelhed* (Holbæk, Nordvestsjælland).

Søgningen gav imidlertid især eksempler på en anden form for kontakt, at København er et sted man tager *ind* til med varer, fisk eller landbrugsprodukter for at sælge dem der, som fx i et citat under *dreng*: *Nogle måneder herefter sejlede drengen til København med en ladning sild* (”blandt fiskere kaldes de drenge til de bliver gift”) (Humblebæk, Nordsjælland). Citaterne om at bringe varer ind til København forekommer ret uforudsigeligt under følgende opslagsord: *dingster*, *distrikt*, *dreng*, *gå*, *have* (V.43.1), *herrefisk*, *klage*, *klods*, *kvase*, *københavnsko*, *københavnslæs*, *last*, *lave*, *levende*, og det ville næppe have været muligt at finde frem til citaterne uden elektronisk søgemulighed.

6. Afslutning

Der er ingen tvivl om at digitaliseringen af bind 1–11 har givet redaktørerne et godt hjælpemiddel ved redaktionen. Men mange andre vil også kunne have glæde af de nye muligheder for at søge på tværs i hele ordbogsmanuskriptet at dømme ud fra de småundersøgelser vi har beskrevet her, og det er noget vi vil tage med os i overvejelserne over hvordan en net-version af ordbogen skal komme til at se ud.

Litteratur

Ordbøger

- Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet* 1–12. 1966–2016. Oslo: Det norske Samlaget.
- ODS = *Ordbog over det danske Sprog* 1–28. Trykt 1919–1956, elektronisk <ordnet.dk/ods>. København: Det Danske Sprog- og Litteraturselskab.
- Ordbok över Finlands svenska folkmål*. Institutet för de inhemska språkens webpublicationer 33. <kaino.kotus.fi/fo>. 2013–. Helsingfors: Svenska litteratursällskapet i Finland.
- ØMO = *Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer* 1– (1992–). København: Center for Dialektforskning, Københavns Universitet & Universitets-Jubilæets danske Samfund.

Anden litteratur

- Gudiksen, Asgerd (2011): Variation i morfologi: 2. svage verbalbøjning i de danske ømål. I: L.-E. Edlund et al. (red.): *Studier i dialektologi och socio-lingvistik: Föredrag vid Nionde nordiska dialektologikonferensen i Uppsala 18–20 augusti 2010*. Uppsala: Gustav Adolfs Akademi, 123–130.
- Gudiksen, Asgerd & Henrik Hovmark (2009): ”Måske husker De noget alle andre har glemt”. I: Asgerd Gudiksen et al. (red.): *Dialektforskning i 100 år*. København: Afdeling for Dialektforskning, Nordisk Forskningsinstitut, Københavns Universitet, 13–64.
- Pedersen, Karen Margrethe (1994): Ordbog og encyklopædi. I: Anna Garde & Pia Jarvad (red.): *Nordiske Studier i Leksikografi 2*. LEDA (Foreningen af Leksikografer i Danmark), 248–256.
- Svensén, Bo (2004): *Handbok i leksikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.

Asgerd Gudiksen
lektor, ph.d.
gudik@hum.ku.dk
Ømålsordbogen

Henrik Hovmark
lektor, ph.d.
hovmark@hum.ku.dk
Ømålsordbogen

Center for Dialektforskning
Institut for Nordiske Studier og Sprogvidenskab
Emil Holms Kanal 2
DK-2300 København S