


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Nettordbøger og Norsk Ordbok – hvordan etablere en vitenskapelig nettordbok	
Forfatter:	Christian-Emil Ore	
Kilde:	Nordiska Studier i Lexikografi 11, 2012, s. 488-499 Rapport från Konferens om lexicografi i Norden, Lund 24.-27. maj 2011	
URL:	http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive	

© Nordisk forening for lexicografi

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.

Nettordbøker og Norsk Ordbok – hvordan etablere en vitenskapelig nettordbok

Christian-Emil Ore

Printed dictionaries have been the way of disseminating systematized information about the use and meaning of words. During the last three decades electronic dictionaries have been common. Most large online dictionaries are based on retro-digitization and are electronic replica of the originals. To defend the cost of converting the printed text to electronic text, the result should be more than an online reproduction. What exactly this 'more' should be is unclear. As a background some dictionaries are examined to show how the digital medium is exploited.

In a paper in 1999 I described a meta-dictionary linking the entries with the source material as a future strategy for NO thus making the dictionary into an electronic scientific publication with reproducible results. In 2002 the meta-dictionary was redesigned to become a pivot in the new combined source database, text corpus and editing system for NO. Vols 5–12 are edited in the new system.

The meta-dictionary and the linked sources have become a valuable repository. The bidirectional linking in the system makes each headword in a source an entry point to NO and the entire system, thus giving dialect users a unique opportunity to see their dialect in the larger context.

The main objective of the NO project is the printed dictionary. However, the system is well prepared to meet the challenge when the 'web' and 'app' age will soon make the online database the main access point.

Nøkkelord: Elektroniske ordbøker, digitalisering, nettpublisering, redigeringsystemer, XML, tekstkoding, vitenskapelighet, formidling

Trykte ordbøker har en lang tradisjon som enerådende teknologi for å formidle systematisert informasjon om bruk og betydning av ord. I løpet av de siste 25 årene har elektronisk publisering blitt mer og mer vanlig. Termer som *nettordbok* og *elektronisk ordbok* blir brukt om slik publisering. Hva som skal til før noe kan sies å være en nettordbok er ennå ikke fastlagt. De skandinaviske språkenes term "ordbok" gir assosiasjoner til et fysisk objekt. Under NFLs 2005 symposium *Elektroniske ordbøker i Norden* var det diskusjon om hvilke krav en skal

sette til en elektronisk tjeneste for at den skal kunne kalles en elektronisk *ordbok*. Termen *ordbok* så ut til å binde flere av deltakerne mentalt til noe som minnet om den tradisjonelle papirordboken. I andre språk brukes termer som er uavhengig av bokformatet, for eksempel *dictionarium*, *glossarium* eller russisk *slovar*. Dette er termer som i en basalbetydning viser til samlinger av ord. I denne artikkelen vil jeg bruke *nettordbok* i vid forstand om alle slags opplysningstjenester for informasjon om ord på nettet og uten noen binding til det tradisjonelle ordboksformatet.

Norsk Ordbok fremstår som en tradisjonell, trykt ordbok. Den er planlagt i 12 bind og skal være ferdig i 2014. Under den synsvinkelen har Norsk Ordbok svært lite med nettordbøker å gjøre. Men bak denne trykte fasaden foregår redigeringen av Norsk Ordbok i et heldigitalt redigeringsverksted. De trykte ordboksbindene er en presentasjon i tradisjonelt trykkformat av bare en del av informasjonen i databasene. Våren 2012 ble de nyredigerte delene av Norsk Ordbok gjort tilgjengelig på internett. Denne første versjonen ligger nær opp til den trykte ordboken slik nettversjonen av *Oxford English Dictionary* (OED) ligger nær opp til sin trykte versjon. I skrivende stund (våren 2012) er det også planer om å publisere alt kildematerialet som finnes i Norsk Ordboks databaser sammen med hovedteksten i ett nettsted. Dette vil være en svært omfattende dokumentasjon av norsk språk diakront og geografisk og ikke minst støtte opp under Norsk Ordboks mål om å gi en uttømmende og vitenskapelig beskrivelse av ordtilfanget i norske dialekter og det nynorske skriftspråket.

I denne artikkelen vil jeg gå gjennom hvordan Norsk Ordboks fremtidige elektroniske publisering passer inn i det større bildet av nettordbøker og andre elektroniske ordbøker. Jeg tar også utgangspunkt i en artikkel jeg skrev for 12 år siden om mulighetene for en alternativ publisering av Norsk Ordbok og dens kilder. Artikkelen het *Metaordboken – et rammeverk for Norsk Ordbok* (Ore 1999) og var basert på et foredrag på konferansen til *Nordisk forening for leksikografi* i Göteborg i 1999. Som nå opplevde en den gangen en rivende teknologisk utvikling. I 1990 foregikk elektronisk publisering på disketter og CD-ROM. På slutten av 1990-tallet var det ikke svært vanlig å publisere ordbøker på nettet eller elektronisk på andre måter. De fleste ordbøkene kom bare i trykte utgaver, noen på CD-ROM og et fåtall som nettordbøker; Bokmålsordboka og Nynorskordboka begge 1991, OED 1996, SAOB 1997, for å nevne noen. Siden 1999 har utviklingen fortsatt. I dag er det irritasjon å spore blant våre brukere av Bokmålsordboka og Nynorskordboka for at disse ordbøkene ikke er tilgjengelig via app-er på smarttelefonene deres.

Da de trykte papirordbøkene var enerådende, handlet en viktig del av leksikografien om å få mest mulig informasjon inn på minst mulig plass, jf. førsteutgaven av Svenséns utmerkede håndbok (Svensén 1987). Det elektroniske

mediet har endret det. En har ubegrenset med plass og store pc-skjermer til å vise teksten. Så alt kan skrives fullt ut. Men i de siste årene har vi så fått smarttelefoner med en relativt sett liten skjerm, og dermed har en igjen behov for et mer stenografisk format. Slik vil det antagelig bølge frem og tilbake. Det eneste som er sikkert, er at det til enhver tid vil være komme nye tekniske løsninger for å slå opp informasjon om ord.

Denne erkjennelsen gjør at det er helt nødvendig for en ordboksredaksjon å foreta en grundig analyse av hvordan informasjonen i ordboken skal organiseres. Her er det selvfølgelig ikke snakk om overflatestruktur som kursiv kontra rettskrift, men om å sette opp en formell modell for hvordan elementene henger sammen. En ordbok skal etter min mening ikke sees som en tekst, men en database, som for eksempel Wordnet (Wordnet). Kunsten er å få informasjonen i denne over til et leselig format – en tekst. Når det gjelder eksisterende trykte ordbøker, er det en pågående diskusjon innenfor humanistisk informatikk om de er tekster eller formelle strukturer (Tasovac 2010).

De trykte ordbøkene fungerer jo utmerket. Det har likevel blitt gjennomført mange såkalte retrodigitaliseringsprosjekter for eldre trykte ordbøker. Målet har vært å gjøre dem om til elektroniske ordbøker, eller nettordbøker om en vil. Inntil relativt nylig var digital lagringsplass svært dyr og internett hadde relativt dårlig overføringshastighet. Den mest farbare veien på 1990-tallet var å skrive av ordbøkene og lagre dem som elektronisk tekst. I løpet av de siste ti årene har datalagring blitt så billig at hvem som helst kan lage digitale faksimiler av boksider. Ordet *digitalisering* har skiftet primærbetydning fra 'konvertering til digital tekst' til 'produksjon av digitale faksimiler'. Det er mye billigere å lage en faksimileutgave av en ordbok enn å taste inn teksten, selv i Kina. Det vesentlige må være hva som er det faktiske formålet med den digitale versjonen. Om en kun ønsker å gi brukerne tilgang gjennom den alfabetiske listen av oppslagsordene, trenger en ikke gjøre annet en å skanne inn sidene, taste inn oppslagsordene og lage lenker fra denne listen til faksimilene. En slik løsning er jo valgt for Otto Kalkars ordbok (KO). Kalkar-løsningen er billig, effektiv og ikke minst tro mot originalen. Dersom ren tilgjengeliggjøring er formålet, kan faksimile-løsningen altså være den beste. Som historikeren Manfred Thaller har formulert det: "A million-page collection with catalogue metadata is, to be precise, considerably more useful than a collection of ten volumes with complete transcriptions encoded according to an elaborate markup scheme developed for the occasion" (Thaller 2001).

Dette er lik tanken bak Google Books, som dessverre ikke har gode katalogdata. Så dersom en ønsker å taste inn hele ordboksteksten, må det være fordi en vil noe mer. Hva dette mer egentlig er, er ikke alltid like klart. Et formål kan være å etablere en elektronisk tekst som skal danne grunnlag for en videre revi-

sjon av ordboken. Det første store eksempelet på dette er digitaliseringen av Oxford English Dictionary (OED) i andre halvpart av 1980-tallet. Siden har slik inntasting blitt alminnelig. Formålet er ikke nødvendigvis elektronisk publisering, men nye trykte utgaver. Men tar en seg råd til å konvertere til elektronisk tekst, har en et hendig og kompakt format med en rekke anvendelsesområder.

De fleste såkalte nettordbøker er i dag basert på elektronisk tekst, enten "born digital" eller retrodigitaliserte trykte utgaver. Det er veldig ujevnt i hvilken grad utgiverne har utnyttet teksten til annet en ren (re)produksjon av "pen" tekst. Mange nettordbøker utnytter ikke den elektroniske teksten til noe mer, og leverer derfor egentlig bare en variant av den enkle "Kalkar"-løsningen til Statsarkivet i Bergen. Hverken Bokmålsordboka, Nynorskordboka eller J. Fritzers norrøne ordbok på nett gir ikke så mye mer enn det. Det samme gjelder for så vidt de 60 ordbøkene i Ordbokshotellet til Norsk Ordbok eller Kunnskapsforlagets "ordnett.no". For de ordbøkene jeg har jobbet med skyldes den enkle nettpresentasjonen flere forhold: manglende kjennskap til mulighetene og tilhørende fravær av ønsker hos redaktørene, manglende ressurser til å dypanalisere den elektroniske teksten og at en rekke av ordbøkene ikke har en stringens som ville gjøre det verdt bryet. Men det finnes eksempler, også i Norden, der en har utnyttet mulighetene på en forbilledlig måte. *Den danske ordbog* har en godt strukturert manuskriptdatabase i bunnen og en har evnet å utnytte den til å lenke inn både ordnett, korpus og interne henvisninger. En her har hatt den fordel at en tenkte strukturert fra begynnelsen (DANLEX 1987).

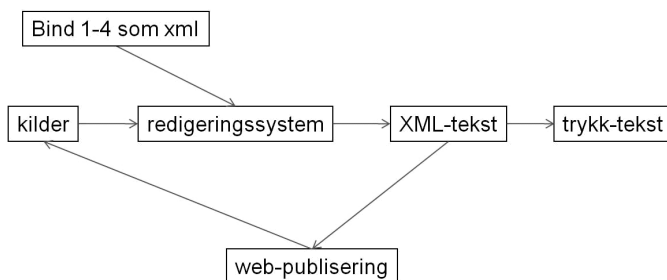
For de såkalte retrodigitaliserte ordbøkene er utgangspunktet mer komplekst. Ordboksredaktørene på 1800-tallet organiserte ikke sine manuskripter etter databasemodeller eller XML-skjema. Men mange eldre ordboksverk er skrevet med en imponerende konsistent artikkelstruktur. Like fullt kreves det et avansert intellektuelt arbeid for å analysere og kode retrodigitalisert ordbøker så langt det går i henhold til en predefinert struktur. Det er ikke uten grunn at TEI Guidelines P5 (TEI P5) for ordbokskoding også har et XML-element "entry-Free" der delementene kan komme i vilkårlig rekkefølge. Da OED ble digitalisert og kodet på slutten av åttitallet, ga en faktisk opp å lage en særlig avansert DTD fordi artiklenes format var for uensartet. Men denne noe flate strukturen har blitt utnyttet godt i den elektroniske utgaven. Den elektroniske OED har en layout som gir assosiasjoner til den originale satsen, men er likevel ganske forskjellig.

En annen stor og viktig retrodigitalisert ordbok er brødrene Grimms *Deutsche Wörterbuch*, *DWB* (DWB). Det digitaliserte manuskriptet ble gitt en detaljert XML-koding i henhold til TEI Guidelines. CD-ROM utgaven fra 2004 bruker kodingen til minste detalj, mens nettversjonen nøyer seg med

strukturkoding i høyre marg og viser originalteksten temmelig kompakt i midtfeltet. Siden er enkel, men ikke elegant som DDOs side. DWB er en del av et tysk ordboksnett www.wortebuchnet.de.

De nordiske svarene på OED og DWB på nettet må sies å være *Ordbok over det danske sprog (ODS)* og *Svenska Akademiens Store Ordbok (SAOB)*. SAOB ble for en stor del konvertert med OCR på nittitallet og kodet og har mange feil, særlig i formatet. Dagens SAOB på nettet hadde kanskje vært bedre om en brukte den elektroniske teksten til å søke i og i tillegg hatt en faksimilevisning av de trykte sidene. Etter sigende skal SAOB skrives inn på nytt slik ODS har blitt det. ODS ble dobbelttastet manuelt i Kina og gitt en enklere XML-koding. Det gjør det i dagens nettutgave mulig å søke på oppslagsord og se enkeltartikler med en delikat layout som minner om den trykte utgaven. Man må kunne anta at utgiverne vil arbeide videre med XML-teksten og knytte den nærmere opp mot DDO og det danske ordnettet.

Norsk Ordbok (NO) er det norske svaret på de store ordbøkene nevnt over. NO er planlagt og designet som en tradisjonell trykt ordbok i 12 bind. Bindene 1 og 2 ble redigert manuelt (skrivemaskin og tradisjonell sats). Manuskriptteksten til bindene 3, 4 og første del av bind 5 ble som en del av redigeringsarbeidet skrevet inn i henhold til et enkelt felt- og kodesystem utviklet på 1980-tallet. Bind 3, 4 og halve bind 5 ble satt elektronisk på grunnlag av den kodete teksten. Det nye redigeringsystemet ble tatt i bruk fra og med bokstaven *i*. Trykksatsen lages i det. NO2014 og EDD har analysert den elektroniske teksten (a-h), gitt den en XML-koding og lastet den inn i redigeringsystemet. Norsk Ordbok redigeres ved hjelp av digital teknologi, men er ennå ikke fullt ut publisert som en digital ordbok, (Ore&Tvedt 2005).



Figur 1 Struktur i Norsk Ordbok-systemet

Selv om Norsk Ordbok ikke har vært planlagt som en digital ordbok og bestillingen fra det norske kulturdepartementet er en trykt ordbok i 12 bind, ligger forholdene godt til rette for å lage en nettversjon av ordboken. I 2011 har det vært arbeidet med å utvikle en slik nettversjon. En prøveversjon ble lansert

våren 2012 (no2014.uio.no). Betrachtingene i den siste delen av denne artikkelen står for min regning og er ikke nødvendigvis NO2014-prosjektets planer eller syn på saken.

Som nevnt ligger manuskriptet til Norsk Ordbok i en database med en struktur som antydet i figur 1. Når teksten skal vise på nettet, skal en da vise det på ODS' maner, det vil si nærmest mulig opp mot det originale satsbildet? Skal en løse opp strukturen og bruke strukturen i XML-teksten for det den er verd? I sistnevnte tilfelle vil vi få en presentasjon av innholdet i Norsk Ordbok som ikke minner så mye om det grafiske utseendet til de trykte bindene. Det er flere forhold en må ta med i betrachtingen når en skal republisere et verk som Norsk Ordbok i nye medier:

1. Norsk Ordbok var og er et prosjekt som skal lage en tradisjonelt utseende ordbok trykt på papir.
2. Redaktørene i Norsk Ordbok er blitt ansatt for å produsere en trykktekst. Kan det for noen medføre at de lager en annen tekst om de skulle ha produsert en nettordbok med dette mediets mange flere muligheter?
3. Er det uetisk eller bryter det med regelverket å lage en nettutgave av Norsk Ordbok som grafisk overhodet ikke minner om originalen? Bryter vi rettighetshaveres ideelle rettigheter tilsvarende Duke Ellingtons versjon av Griegs Peer Gynt suite som ble oppfattet som krenkende av Norsk komponistforening i 1960 og der platen ble forbudt?

På den ene siden er det godt mulig at eldre redaktører vil være fornøyd med et nytt format. På den andre siden kan en vise en viss pietet med hensyn på originalutseende, kanskje ved å gi ut originalversjonen også. Desto eldre et verk en republisere på nett er, desto mindre tenderer en til å tenke over slik hensyn. En god og klar nettpublisering vil øke tilgjengeligheten til innholdet i Norsk Ordbok og dermed bidra til å formidle resultatene av års intellektuelt arbeid. Men det bør nok vurderes om en i tillegg også bør publisere ordboken i et nettformat som svarer til det originale. Men Norsk Ordbok er et prosjekt som har et svært ambisiøst mål:

Norsk Ordbok skal gje ei uttømmande vitskapleg framstilling av ordtilfanget i dei norske dialektane frå 1600 og til i dag og av det nynorske skriftmålet. Hovudvekta skal ligge på grundige tydingsanalysar med klåre definisjonar, kjeldefesta døme på bruk i litteraturen og i målføra, opplysningar om grammatiske eigenskapar og om dialektformer med heimfesting. (www.no2014.uio.no)

Hva som menes med "vitenskapelig" er ikke nærmere beskrevet. Termen "uttømmende" kan implisere at det rett og slett ikke finnes mer informasjon om

ordforrådet enn det som er står Norsk Ordbok. Heldigvis beskrives betydningen av "uttømmende" noe svakere som "tilstrekkeleg, fullgod, dekkjande" i Nynorskordboka og andre norske ordbøker. "Vitskapleg" er mer interessant. "Vitskap" er i følge Nynorskordboka "metodisk innsamling, ordning og etterprøving av kunnskapar og hypotesar etter allment godkjende reglar og krav (om intersubjektiv gildskap [gyldigheit], klarleik osv)". Det er vel stor grad av enighet om at når en beskriver et ords betydninger slik det gjøres i ordbøker eller i leksikalsk semantikk i sin alminnelighet kan det ikke sies å være noen form for hypotesetesting. Det er vel heller i sin beste form en systematisk observasjon og betydningsbeskrivelsene er resultatet av en tilhørende godt begrunnet induksjon. Det vitenskapelige ligger altså i at det er mulig å se det faktiske, empiriske grunnlaget for (induksjons)slutningen. Hvorvidt betydningsbeskrivelsen er skrevet som tekst beregnet for mennesker eller i et mer formelt skjemaspråk er ikke avgjørende for den vitenskapelige gehalten. Et forsøk på å uttrykke betydningen fullstendig analytisk eller i en formell struktur som for eksempel i det generative leksikonformatet er ikke nødvendigvis spesielt vitenskapelig eller forståelig:

One of the several conclusions that may be drawn from this view of lexical meaning [as a set of experiences] is that any attempt to write a completely analytic definition of any common word in natural language is absurd. Experience is far too diverse for that. What a good dictionary offers instead is a typification: the dictionary definition summarizes what the lexicographer finds to be the most common features, in his experience, of the use, context, and collocations of the word. (Hanks 1979)

Jeg vil si det viktigste kravet en kan stille til en vitenskapelig ordbok er at brukerne kan få innsyn i det materialet redaktøren hadde til disposisjon da artikkelen ble skrevet. Det er også en forutsetning at dette grunnlagsmaterialets proveniens som for eksempel kilde, sted og tidfesting er kjent. Det er viktig hvis vi ønsker å kunne trekke pålitelige slutninger fra ordboken med hensyn på utbredelse etc.

Det er ingen tradisjon for at store verkene som ODS, OED, DWB og SAOB skal gi oss mulighet til å etterprøve redaktøren; definitivt ikke i den trykte utgaven, men heller ikke i den digitale utgaven. Litteraturreferansene gir bare kilde for de oppgitte brukseksemplene. Et interessant mellomtilfelle er den delvis redigerte *Ordbog over det norrøne prosaspråk (ONP)*. Nettstedet gir brukeren mulighet til å søke i de ferdig redigerte artiklene (a – em). For resten av alfabetet finnes en database med det digitaliserte kildematerialet og en tilhørende lemmaliste utviklet etter modell av Norsk Ordboks metaordbok. Det er å håpe at redigeringen av ONP blir fullført, og det ville være flott om en kunne lenke opp resten av kildematerialet mot ordboksmanuskriptet. Et uavhengig eksem-

pel er Dictionnaire Étymologique de l'Ancien Français (DEAF) som redigeres i Heidelberg. Her har en digitalisert seddelarkivet og lenket det mot artiklene som redigeres. De eldre delene av ordboken vises som faksimiler av original-sidene. I tillegg er oppslagene i DEAF lenket opp mot tilsvarende prosjekter andre steder i Europa.

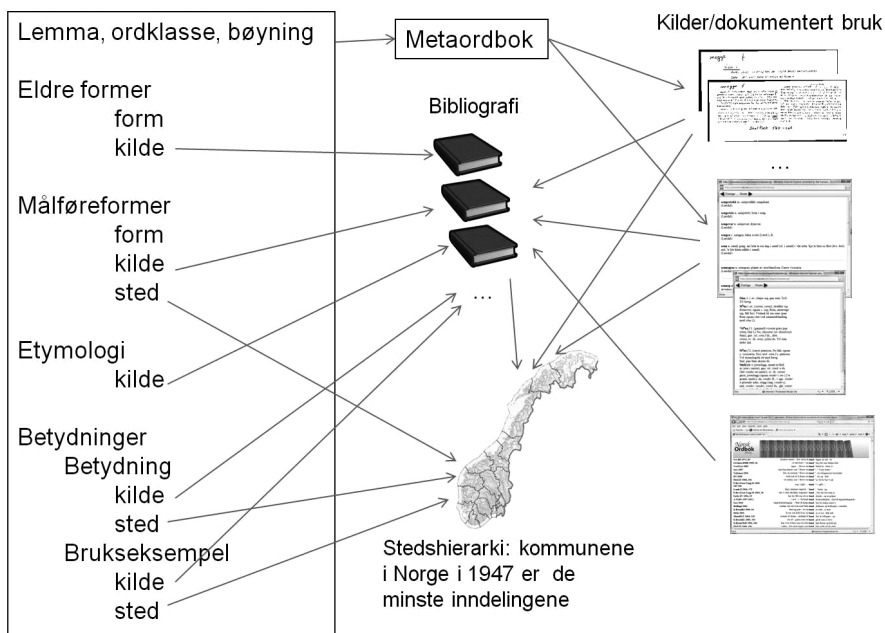
Moderne leksikografiske verktøy som for eksempel Sketch Engine (SkE) eller Tshwanelex inviterer til å lenke korpusanalyse til artiklene i en korpusbasert ordbok. Redigeringsprogrammet Onoma utviklet i 1998–2000 for *Woordeboek van die Afrikaanse Taal (WAT)* av Lexilogik AB er et tidlig eksempel. Redigeringsystemet til Norsk Ordbok (1998 – 2005) er enda et eksempel. Grunnlaget for å lage vitenskapelig etterprøvbare ordbøker er til stede både metodisk og i form av redigeringsprogrammer. Men det har åpenbart ikke blitt gjennomført i noen nettordbøker. Det kan skyldes mange forhold. For mindre, kommersielle ordbøker er det ikke rimelig å forvente noen ambisjoner om vitenskapelighet. Slike ordbøker er praktiske forlagsprodukter og ikke vitenskapelige verk. For de større ordbøkene spiller vel også økonomi inn. Men innen academia er det publikasjonen som teller. Når et vitenskapelig arbeid er fullført og dokumentert gjennom en publikasjon, er det tradisjonelt sett liten vilje til å ta vare på grunnlagsmaterialet. Man kan hevde at det er ingen som kan etterprøve selve innsamlingen og dermed har det ingen verdi å ta vare på det. Det er så, men når grunnlagsmaterialet er borte, kan en heller ikke etterprøve forskerens resonnement fra grunnlagsmaterialet til konklusjonen. Slik sett er resultatene i svært mange empirisk baserte vitenskapelige artikler ikke etterprøvbare og dermed er de i streng vitenskapelig forstand verdiløse.

Norsk Ordboks ambisjon om å være en vitenskapelig ordbok forutsetter dermed at kildematerialet som var tilgjengelig for redaktøren da artikkelen ble redigert, også er tilgjengelig for oss brukere. I en triviell forstand er det tilgjengelig siden seddelsamlingen står i redaksjonens lokaler og alle de andre skriftlige kildene finnes i håndbiblioteket eller i tekstkorpuset. Spørsmålet er om vi kan forlange noe mer? I min artikkel fra Göteborg 1999 viste jeg til eksempler der forskningsprosjekter publiserte det empiriske materialet sammen med de bearbejdede resultatene som en samlet elektronisk utgivelse. At det skulle bli en allmenn trend var i overkant optimistisk. Det har ikke holdt stikk av ulike årsaker, både fordi publikasjonen er det endelige mål og fordi det ikke er noen tradisjon for å dele slike ressurser.

I 1999 hadde Norsk Ordbok 2060 som antatt sluttdato. Det var på det grunnlaget jeg foreslo at modellen med grunnlagsmaterialet og de bearbejdede resultater som en samlet enhet kunne brukes som et substitutt for den ferdige Norsk Ordbok. En slik modell var ikke urealistisk siden mesteparten av bak-

grunns materialet allerede var digitalisert. At dette er blitt modellen også for ONP kan en bare notere seg.

Det er også et annet aspekt med en slik modell brukt for Norsk Ordbok. I den mest puristiske korpusbaserte leksikografien er det bare empiri i form av et tekstkorpus som gjelder. Men for verk som Norsk Ordbok må en også støtte seg på et veldefinert sett av utvalgte kilder i tillegg. Slike er for eksempel spesialiserte fagordbøker og dialektordbøker. En fagordbok gir ofte klare og dekkende definisjoner av (allmenn)ord brukt som tekniske termer. Redaktørene av (gode) dialektordbøker har ofte en nærhet til den lokale tradisjonen og kan vie større plass til de lokale tradisjonsordene, det vil si ord som betegner lokale skikker, redskap, matretter og arbeidsprosesser. I Norge er Njål Veres ordbok *Listamålet* (Vere 1993) med sitt lokale tradisjonsstoff et slikt eksempel. I Danmark har en den mye større *Ømålsordbogen* (ØMO) med et rikt etnografisk materiale. Disse to ordbøkene inkorporerer i stor grad det en med rimelighet kan kalle encyklopedisk informasjon. Norsk Ordbok er en allmennordbok, har begrenset plass og må ikke minst se ordens bruk og i et større landsdekkende perspektiv. Det er derfor naturlig at beskrivelsene en finner i disse spesialiserte ordbøkene kan være klarere og gi mer detaljert informasjon enn det som er mulig i henhold til allmennordboksformatet til Norsk Ordbok.



Figur 2 Sammenhengene i Norsk Ordbok-systemet

Den vitenskapelige ideen bak *metaordboken* er altså at det digitaliserte kildematerialet til Norsk Ordbok kunne bli knyttet til ordboksartiklene i Norsk Ordbok slik at brukerne av systemet kunne studere kildene redaktørene hadde til disposisjon når de skrev ordboksartiklene. Men for den allmenne bruker er det kanskje like viktig at lenkingen gir brukerne tilgang til en rik samling av mindre ordbøker og andre kilder som gir brukerne mer detaljerte opplysninger om for eksempel lokale forhold. Omvendt vil artiklene i Norsk Ordbok bli knyttet til metaordboken på linje med de andre kildene. Koblingen går altså begge veier. Det betyr at en bruker av en digital dialektordbok knyttet til metaordboken kan få se hvilken informasjon redaktørene Norsk Ordbok har redigert for det gitte ordet. Hva mer er; oppslagsordene i dialektordbøkene kan brukes som alternativ søkeinngang til Norsk Ordbok. Det kan ofte være stor avstand mellom de normerte oppslagsordene i Norsk Ordbok og de ulike dialektrealiseringene. Sturla Berg-Olsen og Åse Wetås gjorde rede for muligheter og problemer knyttet til dette i sitt fordrag i Lund (Wetås & Berg-Olsen 2011).

Etter disse generelle betraktningene er det riktig å se litt nærmere på arkitekturen i Norsk Ordbok-systemet for å se i hvilken grad den understøtter en vitenskapelig etterprøvbarehet (se Grønvik og Tvedt 2006 for en beskrivelse av redigeringsystemet). En artikkel i Norsk Ordbok består av tre hoveddeler: oppslaget (lemma, ordklasse og bøyingsinformasjon), en beskrivelsesdel ("hakeparentesen") for eldre kilder, avvikende skriftformer, dialektrealiseringer og etymologi, samt den sedvanlige artikkeldelen med betydningsbeskrivelser og brukseksempler. Informasjonen i oppslagsdelen er i henhold til 1938-rettskrivningen og er normativ. Den trenger derfor ingen begrunnelse. De ulike delene av hakeparentesen vil på trykk kunne bestå av ordformer med tilhørende kilde- og stedfesting. Betydningsdelen kan være en enkel liste eller et hierarki; informasjonsteknisk sett kan vi kalle den en trestruktur. Markerte betydninger tradisjonelt angitt i trykk med tall eller bokstaver svarer til en node i et slikt tre. Til hver node finnes det i databasen en informasjonspakke som kan inneholde definisjonstekst, krysshenvisninger, redaksjonelle brukseksempler og sitater fra kildematerialet med mer. Definisjoner og brukseksempler kan ha tilknyttet sted og kildefesting, mens sitatene selvfølgelig må ha det. Stedfestingen er oftest en generalisering til et større område basert på de stedene som er angitt i kildematerialet. Henvisningene til de litterære kildene og sitatene, vil være en delmengde av det som finnes i kildematerialet. Utvalget og generaliseringene er et resultat av redaktørens intellektuelle arbeid. Dette er standard for deskriptive ordbøker som Norsk Ordbok. Men det gir altså ingen direkte tilgang til bakgrunns materialet. En standard digital utgave av Norsk Ordbok vil ikke endre på det. En slik digital utgave bare kunne hjelpe brukeren til å skjønne de stenografiske steds- og litteraturhenvisningene og eventuelt også gjøre det mulig å søke

geografisk eller bibliografisk. For Norsk Ordbok er listen over de 6 000 litterære kildene lagret i en standard bibliografisk format (MARC). For stedfestingen er norske kommuner i 1947 den høyeste nøyaktigheten, og er byggeklossene i stedfestingene i ordboken. Kommunegrensene er blitt digitaliserte, og det åpner for kartbasert søk og visning (se Grønvik og Holmen 2011).

Norsk Ordbok har i tillegg kobling til kildene via metaordboken. Hver artikkel er dermed knyttet til de kildene som finnes for oppslagsordet. Redigering gjør det også mulig å koble kilder, som sedler, andre ordbøker og resultatet av korpussøk direkte til (del)betydningene og til de tre kategoriene eldre kjelder, avvikende skriftformer og målføreformer. Dette er ikke brukt i så stor grad på grunn av høyt arbeidspress. Så en må nøye seg med å sjekke det samlede grunnlagsmaterialet for hele artikkelen. Som nevnt over, kan resultatet av korpussøk lagres i metaordboken i form av kwic-konkordanser i metaordboken. Dermed kan en sjekke hvilket utvalg fra korpus som er brukt. Denne mekanismen er heller ikke brukt i noen særlig grad. Det kan også skyldes at det er tungt å redigere ned en resultatkonkordanse til de linjene en faktisk har brukt. Men dersom noe fra korpus brukes som sitater må det lages sedler som lenkes i metaordboken. Norsk Ordbok bruker IMS Corpus Workbench (CW) som korpusverktøy. Det har en relativt enkel måte å vise tilslag. En kunne i fremtiden tenke seg å bruke for eksempel Sketch Engine (SkE) som leverer ferdigere analyser av korpusmaterialet. Disse skissene kan lett lagres i metordboksystemet for senere sjekk. Denne lille gjennomgangen viser at det er om ikke huller i systemet for å etterprøve redaktørene, så er det visse svakheter. Men om en går tilbake til forslaget til metaordbok fra Göteborg 1999 har en kommet veldig langt mot en fullstendig realisering av en vitenskapelig etterprøvbar ordbok.

Den første versjonen av Norsk ordbok som nettordbok, *Norsk nettordbok*, ble publisert våren 2012 og er en tradisjonell nettordbok med funksjonalitet som OED eller DWB. Den vil i løpet av 2012 etter hvert med kartfunksjonalitet. Det neste skrittet er å integrere kildematerialet på en god måte. Det er en utfordrende oppgave å formidle et slikt kompleks med hensyn på klarhet for den vanlige bruker. Men det vil gi en enestående inngang til kunnskap om norsk språk og synliggjøre den formidable innsatsen til redaktørene av Norsk Ordbok.

LITTERATUR

DANLEX = Descriptive tools for electronic processing of dictionary data: studies in computational lexicography. The DANLEX-group, Ebba Hjort (et al.). *Lexicographica*, Series Maior 1987. Tübingen: Niemeyer.

- DEAF = Dictionnaire Étymologique de l'Ancien Français, www.deaf-page.de (sjekket 05-11-2011)
- DWB = Deutsche Wörterbuch www.dwb.uni-trier.de (sjekket 05-11-2011)
- CW IMS = www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/ (sjekket 05-11-2011)
- Grønvik, Oddrun og Tvedt, Lars Jørgen, 2006: Norsk Ordbok 2014 – presentasjon av eit komplekst leksikografisk verktøy. I: Nordiske Studier i Leksikografi nr 8, København.
- Holmen, Jon og Grønvik, Oddrun, 2012: Norsk dialektatlas i digital utgåve og språkdata på kart. I: Nordiska studier i lexicografi 11. Lund. (Skrifter utgivna av Nordiska föreningen för lexicografi 12.)
- Hanks, Peter, 1979: To what extent does a dictionary definition define? I: R. R. K. Hartmann (ed.), Papers from the 1978 B.A.A.L. Seminar on Lexicography. Exeter Papers in Lexicography.
- OED = Oxford English Dictionary: www.oed.com/ (sjekket 05-11-2011)
- Ore, Christian-Emil, 1999: Metaordboken – et rammeverk for Norsk Ordbok. I: Nordiska studier i lexicografi 5. Göteborg.
- Ore, Christian-Emil, Tvedt, Lars Jørgen, 2006: Digital sats eller digital satsing. I: Nordiske Studier i Leksikografi nr 8. København.
- SkE Sketch Engine: www.sketchengine.co.uk (sjekket 05-11-2011)
- KO = Kalkar, Otto, 1881–1918: Ordbok til det ældre danske sprog (1300-1700). www.hist.uib.no/kalkar/
- Svensén, Bo, 1987: Handbok i lexicografi : principer och metoder i ordboksarbetet. Esselte studium; Tekniska nomenklaturcentralen. Stockholm.
- Tasovac, Toma, 2010: Reimagining the Dictionary, or Why Lexicography Needs Digital Humanities, Digital Humanities 2010, Kings College, London, UK
- TEI P5 = Text Encoding Initiative P5: www.tei-c.org (sjekket 05-11-2011)
- Thaller, Manfred, 2001: From the Digitized to the Digital Library, D-Lib Magazine, February 2001, Volume 7 Number 2, ISSN 1082-9873.
- Tshwanelex: tshwanelex.com/tshwanelex (sjekket 05-11-2011)
- Vere, Njål, 1993-1995: Grammatikk, ordbok, stadnamn og folkeminne I-III, Vanse.
- Wetås, Åse og Berg-Olsen, Sturla, 2012: Norsk Ordbok på nett: Utfordringer og løysingar . I: Nordiska studier i lexicografi 11. Lund (Skrifter utgivna av Nordiska föreningen för lexicografi 12.)
- WAT = Woordenboek van die Afrikaanse Taal. <http://www.wat.co.za> (sjekket 05-11-2011)
- Wordnet = wordnet.princeton.edu (sjekket 05-11-2011)
- ØMO = dialekt.ku.dk/dialekter/oemaalsordbogen/ (sjekket 05-11-2011)

Christian-Emil Smith Ore

Institutt for lingvistikk og nordiske studier, Det humanistiske fakultet, Universitetet i Oslo.

c.e.s.ore@iln.uio.no