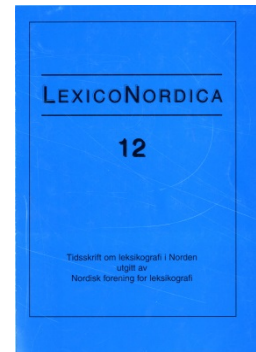


# LexicoNordica

Titel: Norsk Ordbok – også ei elektronisk ordbok?  
Forfatter: Kristin Bakken & Christian-Emil Ore  
Kilde: LexicoNordica 12, 2005, s. 7-19  
URL: <http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive>



© LexicoNordica og forfatterne

## 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 LexicoNordica (1-16) 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.

*Kristin Bakken & Christian-Emil Ore*

## ***Norsk Ordbok – også ei elektronisk ordbok?***

The paper gives a short presentation of *Norsk Ordbok* (The Norwegian dictionary) as to scope and history. The recent computerization efforts have given the project a database with the data structured in a way that invites exploitation in the format of an electronic dictionary. The authors account for the information categories that are reflected in the new database, and on how these categories can be activated by the search system developed to access information from the database. It is argued that the database and the search system already constitutes an electronic dictionary, with the defined user group restricted to the editors for now. Made public, the authors think that the electronic dictionary both will obtain new users, and will be put to other uses than the paper version. In conclusion three focuses for future work are outlined. Firstly the project wants to present the dictionary integrated with its sources. Secondly, it is discussed how the user interface could accommodate different user groups, and finally it is commented on how a well designed cross-reference system may be exploited for new kinds of searches.

### **1. Innleiing**

*Norsk Ordbok* er ei historisk dokumentasjonsordbok som både har det nynorske skriftspråket og dei norske dialektane som målspråk. Ordboka er altså ei kombinert dialekt- og skriftspråksordbok. Den kronologiske avgrensinga er ikkje den same for dialektmaterialet som for nynorsken – naturleg nok – sidan nynorsken oppstod som skriftspråk ca. 1850, medan dei eldste dialektoppskriftene for norsk vart gjorde på 1600-talet. Ordboka skal difor dokumentere dei norske dialektane frå 1600-talet til i dag, og nynorsken frå 1850 til i dag. Oppslagsorda i ordboka er normert nynorsk, og dialektformene blir førde til det tilsvarende nynorskoppslaget. I dei tilfella det ikkje finst normerte nynorskord som svarar til dialektordet, inneber normeringa av oppslagsordet at nynorskleksikonet blir utvida med eitt ord.

Dette doble fokuset gjev ein etter måten kompleks artikkelstruktur. Artikkelen skal både dokumentere ordhistorie, tydingsstruktur vist ved kjeldefesta døme, ordgeografisk utbreiing, og formelle variantar. Grovt sett inneheld ein artikkel i *Norsk Ordbok* desse informasjonskategoriane: oppslag, grammatikk, eldste kjelder, dialektvariantar med heimfesting, etymologi, tydingsstruktur med definisjonar og kjeldefesta døme, samt leksikalisererte frasar som har oppslaget som kjerneord. Dette gjev artiklar som kan variere frå éi line til åtte-ti sider.

*Norsk Ordbok* er ei gammal ordbok; arbeidet tok til i 1930. Men redaksjonen vart omorganisert i 2002, etter at ein liten redaksjon hadde gjort fire band ferdige. Frå det året vart redaksjonen organisert som eit eige prosjekt, og det vart gjort mange nyttilsetjingar. Band 5 kom i januar 2005, og innan 2014 skal heile ordboka på 12 band vere fullført.

Friske løyvingar frå Stortinget vart følgde av nye strenge krav til redaksjonen. Kravet om fullføring til 2014 er alt nemnt, men nokre føresetnader for å greie ei slik fullføring vart også formulerte som vilkår for det nye prosjektet. Redaksjonen måtte organiserast på nytt med sterkare fokus på målstyring, det redaksjonelle arbeidet måtte effektiviserast, og sist, men ikkje minst, fekk vi krav om i større grad enn før å ta i bruk digitale verktøy. Digitalisering og it-baserte redaksjonelle rutinar vart sett på som ein føresetnad for å greie å effektivisere arbeidet.

Oppdraget til redaksjonen var heilt klart å fullføre ei tolvbands papirutgåve. Digitaliseringskravet var fyrst og fremst knytt til effektiviseringsbehovet, men det var også motivert av at Dokumentasjonsprosjektet på 1990-talet hadde digitalisert setelarkivet til ordboka, og slik var det ei utfordring å hauste tids- og kvalitetsgevinstar av dette store arbeidet. Ei elektronisk utgåve av ordboka var ikkje noko tema i utgreiingane som låg til grunn for omorganiseringa av ordboksredaksjonen. Men etter snart tre år med ny verksemd er stoda slik at det faktisk ligg føre ei elektronisk utgåve av ordboka som biprodukt av det generelle digitaliseringsarbeidet. Dette opnar opp nye og interessante perspektiv for *Norsk Ordbok*, og i denne artikkelen vil vi presentere og kommentere nokre slike perspektiv.

Her må det nemnast at vi, artikkelforfattarane, representerer to ulike synsstader, sidan den eine, Christian-Emil Ore frå Eining for digital dokumentasjon, står ansvarleg for alt IT-basert utviklingsarbeid for *Norsk Ordbok* sin redaksjon, medan den andre, Kristin Bakken, er fagleg ansvarleg for ordboksredaksjonen. I planlegginga av ei framtidig elektronisk utgåve, er vi avhengige av nær dialog, og slik dialog er då også denne artikkelen er eit resultat av.

## **2. *Norsk Ordbok* – frå tekst til database**

Då arbeidet med å lage redigeringsdatabasen tok til, var dei fire første banda alt publiserte, og redaksjonen var godt i gang med band 5. Å utvikle ein redigeringsapplikasjon føreset normalt at ein har formulert veldefinerte redigerings- og formatreglar på førehand. Redigeringsreglane for *Norsk Ordbok* var i utgangspunktet baserte på den vanlege

standarden for slike historisk orienterte ordbøker frå første halvparten av 1900-talet. Men redigeringspraksisen hadde utvikla seg over tid, og den redigeringshandboka som redaktørane hadde å hjelpe seg med, var altfor tynn og underspesifisert. I 2000 var difor mykje av redigeringspraksisen basert på munnleg traderte konvensjonar som ikkje var nedfelte i skriftleg form. Eit klart formulert mål då vi gjekk i gang med å utvikle eit elektronisk redigeringsverktøy, var difor å stramme inn og formalisere redigeringsreglane. Ikkje berre var dette naudsynt av omsyn til det elektroniske utviklingsarbeidet, vi meinte også at det ville gjere opplæringa av nye redaktørar lettare og sjølve redigeringsprosessen meir effektiv. Utfordringa var å lage ein database som både skulle innehalde dei eksisterande banda, som hadde blitt redigerte frå 1950 til 2002, og som skulle kunne vere eit funksjonelt verktøy for arbeidet med resten av ordboka.

Utviklingsarbeidet tok utgangspunkt i dei eksisterande banda, i handboka og i den oppfatninga redaktørane hadde av gjeldande praksis. Prosessen varte i nesten to år. Eit viktig mål var å bryte opp artiklane i ein veldefinert hierarkisk og rekursiv trestruktur. Vi freista å unngå at dei minste einingane som kjem til uttrykk som felt i innskrivings-skjemaet, skulle ha ein indre struktur. Vi ville likeins unngå å lage eit system der redaktørane skrev fritekst med kodar (XML-taggar) inne i teksten i eit felt. Vi har på det næraste greidd det. Dermed kan ein ha førehandsdefinerte menyar for det meste av det faste redaksjonelle språket. (Sjå elles Grønvik 2005a og Grønvik 2005b for nærare utgreiing.)

Det stringente formatet opnar for eit utal moglege søk (sjå punkt 3 nedanfor). Databasen er sjølv sagt ei samling med artiklar som beskriv orda i språket. Men ein kan også snu på det og seie at databasen inneheld eit sett med meiningar (definisjonane) som er indekserte ved hjelp av t.d. oppslagsord, geografiopplysningar, bruksområde og stilistisk informasjon.

Formatet gjer det enkelt å skrive kvar artikkel som eit XML-dokument, men vi har valt å implementere redigeringsapplikasjonen i ein relasjonsdatabase (ORACLE). Eit slikt databasesystem gjev (førebels) ein betre dataintegritet, som det er lettare å oppnå slik enn ved eksisterande reine XML-databasar. Reine XML-funksjonar som generering av sats og formatsjekking tek vi oss av ved å byggje opp XML-dokument med utgangspunkt i tabellstrukturen i databasen. Relasjonsdatabasen gjer det også enklare å kople saman manuskriptdelen med alle kjeldedatabasane og metaordboka (sjå punkt 5.1 nedanfor).

I arbeidet med redigeringsapplikasjonen tok vi altså utgangspunkt i dei fire eksisterande banda av *Norsk Ordbok*. Sidan praksisen viste seg

å vere variabel, blei det som nemnt naudsynt å stramme inn og spesifisere redigeringsreglane. I nokre tilfelle førde dette revisjonsarbeidet til regelendringar. Sjølv om vi vil understreke at dette arbeidet inneber eit vesentleg kvalitativt løft for ordboka, representerer dei endringane som vart gjorde mellom band 4 og 5, også ei utfordring for databaseutforminga. For å sikre konsistens i heile ordboka og for at den elektroniske versjonen også skal vere komplett, er det viktig og naudsynt å få lagt inn dei fire første banda i databasen. For å greie dette har vi laga eit analyseprogram som har teke utgangspunkt i satsbiletet (halvfeit, kursiv og liknande), og konvertert den inntasta teksten til eit XML-format som svarar til strukturen dei nye artiklane i databasen har. Resultatet så langt er bra, og vi har no greidd å få inn dei fire første banda analyserte i hovudstrukturar. På mikronivå står det att noko arbeid som vi har sett av hausten 2005 til å få unna. Mykje av dette arbeidet vil bli gjort på systemutviklingssida, men noko tilpassing til databasestrukturen vil truleg rett og slett krevje redaksjonelt basert revisjon. Dermed vil også denne ferdigstillinga skje i nær dialog mellom systemutviklarane og redaksjonen.

### 3. Søkjemoglegheiter i databasen

Det nye databaseformatet for *Norsk Ordbok* er altså ein realitet, og med nokre unntak er band 1–5 no tilgjengelege via databasen. Saman med søkjesystemet som også er på plass i ein første versjon, gjer dette at *Norsk Ordbok* allereie er ei elektronisk ordbok, sjølv om ho førebels berre blir brukt av redaksjonen internt, og sjølv om søkjesystemet enno ikkje er utvikla med tanke på nærare definerte brukargrupper. Vi vil først seie litt om kva slags søk databaseformatet no tillèt. Desse felte er blant anna no søkbare:

- Lemma
- Ordklasse
- Markert bøyingsklasse
- Visse uttaletrekk
- Etymologi
- Kjeldereferansar (kodifiserte)
- Geografiske heimfestingar (kodifiserte)
- Bruksmarkeringar, dvs. opplysningar om stilnivå, fagfelt, ekstensjonsavgrensing osv. (kodifiserte)
- Definisjonar
- Sitat

- Fleirordige oppslag

Felta har litt ulik type fyll. Nokre av dei vert fylte ved å gjere val i ferdigdefinerte menyar. Døme på dette er felta for ordklasse, bøyingsklasse, etymologi, kjeldereferansar, geografiske heimfestingar og bruksmarkeringar. Andre felt er fritekstfelt, og det gjeld typisk nok sjølvle lemme, definisjonane, sitata og dei fleirordige oppslaga. I dei kodifiserte felta kan ein då søkje ved hjelp av lukka menyar, medan ein søker på einskildord i alle fritekstfelt. Alle søk kan sjølvstyktrunkast, og dei kan gjerast kombinatoriske med *og*. Det førebelse søkjesystemet opnar for å utforme søkjetrea i fleire variantar, og slik tilpasse det til individuell bruk. Døme på søk som det er kurant å gjere alt no, er å hente fram alle maskuline substantiv som innheld definisjonar der *mann* førekjem, eller alle verb med e-bøying som har fransk opphav, eller ord som er heimfesta til Finnmark og som har samisk opphav. Vi kan også finne fram til alle ord med nedsetjande stilverdi, eller alle plantenemningar (som høyrer til fagfeltet botanikk).

For redaksjonen har dette søkjesystemet blitt eit uvurderleg verktøy, og vi brukar det for å sikre konsistent redigering på tvers av variablar som redaksjonstid (kan variere frå 1950 til i dag), ulike redaktørar (vi har pr i dag 21 redaktørar i aktivitet) eller redaksjonsstad (i tillegg til redaksjonen i Oslo har vi ein filial i Trondheim). Vi har også fått hjelp til å halde orden på det interne krysstilvisingssystemet, og vi brukar søkjesystemet til å skaffe fram empirisk underlag for å gjere justeringar i redigeringshandboka og av redigeringspraksisen. For redaksjonen er altså søkjesystemet blitt eit kvalitetshevingsverktøy, i tillegg til at sjølvle oppslagsformatet er blitt svært effektivt. Så langt har vi ikkje lagt ned arbeid i å tilpasse databasen og søkjesystemet ekstern bruk, men dette er ei oppgåve vi ønskjer å kome i gang med i år, og denne artikkelen er slik eit første spadestik i retning av å planleggje og realisere ei elektronisk utgåve av *Norsk Ordbok*.

#### **4. Føresetnader for ei elektronisk utgåve av *Norsk Ordbok***

De Schryver (2003) siterer Cerquiglioni som reknar med at datamaskinbasert leksikografi kan vurderast i høve til tre nivå. På det lågaste nivået brukar leksikografane datamaskina til hjelp for å framstille papirordbøker. Den typiske situasjonen er at ein brukar eit tekstbehandlings-program til å utarbeide ordboksmanuskript. Delar av band 3 og band 4 av *Norsk Ordbok* vart framstilte på denne måten. Det andre nivået Cerquiglioni viser til, er å overføre eksisterande papir-

ordbøker til eit elektronisk medium. Nettutgåva av *Svenska akademis ordbok* er ei slik digitalisert papirordbok (sjå Cederholm & Rogström 2001). Det tredje og mest avanserte nivået i datamaskinbasert leksikografi, er å konseptualisere ordbøker direkte for det elektroniske mediet. I dei siste åra har ein sett stadig meir kreativ utnytting av det elektroniske mediet for å lage ordbøker på nye måtar. Frå leksikografikonferansane dei siste åra har det blitt demonstrert stadig nye konsept, det kan vere tospråklege ordbøker integrert med interaktive læringsopplegg, ordbøker med fyldig ensyklopedisk utfylling i form av bilete, film og lyd eller dynamiske ordbøker som kan tilpassast individuelle brukarar. (Interessante døme er Kruijsen & Swanenberg 2004, Løfberg et al. 2004, Pelissier & Becker 2005, Gawinski under trykking frå den nordiske leksikografikonferansen i Sønderborg 2005.) De Schryver 2003 gjev eit omfattande oversyn over korleis ein kan fornye og vidareutvikle ordboka som genre og dermed også ordboksbruken innanfor det elektroniske mediet.

Kvar vil vi plassere den elektroniske utgåva av *Norsk Ordbok* (heretter e-NO) i eit slikt bilete? På den eine sida vil e-NO sjølvsagt ha eit motsvar i papirutgåva, for mandatet vårt vil vere uendra; vi skal fullføre ei papirutgåve. Dei informasjonskategoriane som papirutgåva skal ha med, vil difor også kjenneteikne e-utgåva. På den andre sida er jo desse informasjonskategoriane no lagde inn i ein strukturert og fleksibel database, og dette formatet gjer det mogleg å byggje opp ei e-utgåve som er spesialtilpassa for det elektroniske mediet. I denne samanhengen må det understrekast at databasen er bygd opp med tanke på menneskeleg bruk og brukarar, dvs. at det ikkje er teke høgd for å gjere denne leksikalske databasen maskinleseleg. Det må likevel nemnast at databaseformatet gjev eit stringent utgangspunkt for eventuell vidareutvikling i slik lei. Men i fyrste omgang ønskjer vi no å lage ei e-ordbok for menneske ved å utnytte dei moglegheitene som databasen gjev oss, og ved å stille oss ganske fritt til i høve til strukturen i papirutgåva. Dermed er nok ambisjonen vår å realisere ei elektronisk ordbok på det elektroniske mediet sine premissar. Men vi må presisere at vi er bundne av det faktum at databasen i utgangspunktet er bygd opp på grunnlag av informasjonskategoriane i ei etablert papirordbok.

Før vi held fram med å trekkje opp nokre ideal for e-NO, må det også seiast at forlagsrettane til papirutgåva av *Norsk Ordbok* ligg til Det norske Samlaget. Vi har sett i gang ein prosess med forlaget for å få avklart kva for vilkår som må vere oppfylte for å gjere e-NO tilgjengeleg på internett. Førebels er dette ikkje formelt avklart, men frå vår side er det heilt klart at målet må vere å gjere e-NO gratis

tilgjengeleg over internett. Sidan NO 2014 er eit svært kostbart prosjekt som heilt ut er kosta av det offentlege, dvs. av skattebetalarane, er det etter vårt syn urimeleg å be brukarane betale ein gong til for å få tilgjenge til e-utgåva av ordboka.

Prisen på prosjektet NO 2014 er éin av grunnane til at vi ønskjer å realisere ei elektronisk utgåva av ordboka. Den norske staten gjer eit tungt kulturøkonomisk løft ved å fullfinansiere *Norsk Ordbok* i tida fram til 2014. Sjølv om dette førebels ikkje er eksplisitt uttalt frå politisk hald, må det vere i tråd med Stortinget sin intensjon å gjere *Norsk Ordbok* tilgjengeleg for så mange som mogleg. Det er klart at ei elektronisk utgåve vil nå vesentleg mange fleire enn den relativt kostbare papirutgåva. Men talet på brukarar er berre éin av grunnane til at e-NO er ønskjeleg frå vår synsstad. Nye brukarar heng nemleg også saman med ny bruk av ordboka, og denne nye bruken er fagleg sett svært interessant for oss.

Tradisjonell bruk av papirordbøker er bunden av alfabetet og har eit typisk partikulært fokus. Ein slår opp ord ein i utgangspunktet har høyrte før, og ein leitar etter dei på rett plass i alfabetet. I og med e-utgåva slepp ein både den alfabetiske tvangstrøya og det einsidige partikulære fokuset. Døme på ny bruk som e-utgåvene opnar opp for, er at ein kan variere søkjeinngangane, dvs. at ein ikkje lenger er avhengig av å gå inn i ordartiklane via oppslagsordet. Dermed treng ein ikkje lenger kjenne til ordet ein leitar etter, for ein kan gjere ”fuzzy” eller kombinatoriske søk, slik at ein kan gå frå eit upresist utgangspunkt og finne presise opplysningar. (Sjå elles nedanfor under punkt 5.2.) Det er også viktig at ein i ei elektronisk ordbok kan gjere systematiske søk. Det vil vere mogleg å leite etter mønster heller enn einskildord, for ein kan til dømes hente fram alle ord som deler visse eigenskapar. I tillegg vil ei elektronisk ordbok kunne gje direkte koplingar til kjeldene ordboka byggjer på (sjå nedanfor), det elektroniske mediet gjer det lettare å få tilbakemeldingar og å gjere revisjonar, og til sist kan mediostrukturen eller det interne lenkesystemet (krysstilvisingane) aktiviserast på meir systematiske måtar også i søkjesamanhengar. Alt dette opnar opp svært interessante leksikografifaglege perspektiv for oss, og det motiverer for å arbeide fram e-NO.

## **5. Visjonar for ei elektronisk utgåve av *Norsk Ordbok***

På bakgrunn av føresetnadene spesifiserte under punkt 4 ovanfor, og med det utgangspunktet som er lagt i og med databasen (jf. punkt 2), vil vi no skissere nokre visjonar for e-NO. Vi vil i denne samanhengen vise



til de Schryver (2003) for inspirerende perspektiv. Her vil vi konsentrere oss om tre moment:

1. Integrering av interne og eksterne kjelder
2. Gjennomtenkt søkjegrensesnitt med tilpassing for ulike brukargrupper
3. Utnytting av mediostrukturen

### *5.1. Integrering av interne og eksterne kjelder*

Den fyrste visjonen vi har for e-NO burde ikkje vere vanskeleg å realisere, sidan vi alt har etablert ei redaksjonell plattform der kjeldene og ordboksdatatabasen er kopla saman. Det er dette vi viser til når vi her refererer til "interne kjelder". Eining for digital dokumentasjon har laga eit elektronisk register over alt digitalisert tilfang som det er relevant å redigere *Norsk Ordbok* på grunnlag av (metaordboka). Mellom metaordboka og ordboksdatatabasen er det etablert koplingar som gjer det mogleg for brukarar å få tilgjenge til den empirien som ordboksartikkelen byggjer på. Slår ein altså opp eit ord i ordboksdatatabasen, kan ein gå vidare for å sjå redaktøren i korta. I tillegg til alt tilfang som er indeksert i metaordboka, har vi også etablert eit tekstkorpus, og vi er nett no i ferd med å ta i bruk eit system der relevante korpustreff også blir lenka til ordartiklane i datatabasen. Til saman har vi difor allereie etablert ei interessant vidareutvikling av omgrepet dokumentasjonsordbok. Det er jo berre ein brøkdel av dokumentasjonen som faktisk får plass i ordartikkelen som døme, men e-NO opnar for at tilleggsdokumentasjon kan bli tilgjengeleg via lenker frå den elektroniske ordartikkelen. Dette gjer ordboka etterprøvbar på ein heilt ny måte, og vi meiner difor at e-NO med større rett enn papirutgåva vil kunne klassifiserast som ei vitskapleg dokumentasjonsordbok. I e-NO vil ordartikkelen vere ei opning inn mot eit stort relevant språkleg tilfang.

Når det gjeld eksterne kjelder, er vi førebels berre på idéstadiet. Men Eining for digital dokumentasjon har alt etablert fleire ensyklopediske databasar på same plattform som språkdatatabasen til NO. Gjenstandsdatatabasar for musea, bilettdatabasar og kartdatabasar finst allereie, og det er klart at det her er store moglegheiter for å kople NO-databasen til basar som innheld bilete, lyd, kart eller ensyklopedisk kunnskap. Når det ikkje lenger finst fysiske plassbegrensingar, er det ingen grunn til at NO t.d. ikkje skulle kunne (del-)illustrerast, og som dialektordbok kunne vi absolutt vere interesserte i utbreiingskart og

uttaleprøver. (Jf. Kruijsen & Swanenberg 2004 om slikt tilleggsmateriale i dialektordbøker frå Nederland og Belgia.) Dei faglege og tekniske moglegheitene er store; kva som kan realiserast vil avhenge av ressursituasjonen.

## *5.2. Gjennomtenkt søkjegrensesnitt med tilpassing for ulike brukargrupper*

Som alt nemnt har vi alt svært gode søkjemoglegheiter inn i databasen i dag. Men vi har enno ikkje teke stilling til kva for søk vi særleg skal leggje til rette for. Dette heng saman med spørsmålet om kven brukarane av ei framtidig e-NO er, og kva vi ønskjer at desse brukarane skal bruke ordboka til. Eit hovudpoeng må vere at e-NO både skal kunne brukast på same måte som papirutgåva, dvs. til å slå opp lemma og lese heile eller delar av ordartikkelen, og desutan at e-NO skal kunne brukast på nye måtar, dvs. til å gjere systematiske søk på tvers, og til å søkje ved hjelp av ulike og kombinerte kriterium både på mikro- og makrostrukturnivå.

Dei nye brukarane tenkjer vi oss kunne vere alle som treng spesialtilpassa dokumentasjon som underlag for ulikt språkarbeid. Det kan vere som underlag for språknormering, språkforskning, språkdidaktikk og profesjonell språkproduksjon. For alle desse vil e-NO vere eit meir effektivt hjelpemiddel enn papirutgåva. Vi ønskjer dessutan også at den mangfaldige språkinteresserte ålmenta som finst i Noreg – og utanfor – vil ha glede og nytte av det mangfaldet av tilgjengeleg informasjon som vil finnast i e-NO.

Vi har alt vore inne på korleis e-NO opnar opp for meir systematisk bruk av ordboka enn før. Dette burde gjere e-NO meir interessant for språkforskarar enn det har vist seg at papirutgåva har vore. Fleire har peikt på at elektroniske utgåver faktisk aukar verdien av ordbøker som forskingsverktrøy, jf. t.d. Ruus (2002). Vi tykkjer difor det er verdt å seie litt meir om potensiell bruk av e-NO for framtidig språkforskning.

Ein grunnleggjande føresetnad for språkvitskapleg forskning er at språket faktisk er regelmessig eller prega av repeterte mønster. Språkforskning handlar om å beskrive slike mønster eller regelmessigheiter og å uttrykkje dei i ulike regelformat, og dessutan er ein oppteken av å forklare regelmessigheitene. Praktisk leksikografi kviler på den eine sida på språkvitskapleg forskning, for analysen av einskildord byggjer på tidlegare forskarar sine regelformuleringar og forklaringar. På den andre sida burde leksikografane ideelt sett også kunne levere foredla data til dei forskarane som er opptekne av å kartleggje dei

mønsterbetinga aspekta av språkbruken. Tradisjonelt har likevel dei store dokumentasjonsordbøkene vore vanskelege å utnytte til slikt arbeid, nettopp fordi fokuset som ein funksjon av papirordboksgenren har vore partikulært. Med dei elektroniske ordbøkene er dette endra, og vi meiner at e-NO på ein heilt annan måte enn papirutgåva vil kunne brukast til å kartleggje norske ordlagingsmønster, ordgeografiske mønster, fraseologi og kollokasjonsoppbygging, mønster som har med omgrepsdanning og semantisk struktur å gjere, kartlegging av leksikalske lån og tilpassing av framandt ordstoff og dessutan fonologiske dialektbetinga variasjonsmønster. Integreringa av kjeldene er i denne samanhengen uunnverleg, for forskarane kan sjølve gå tilbake for å vurdere om dei fortolka, og slik sett foredla, opplysningane som blir presenterte i ordartikkelen, er lingvistisk forsvarlege. Vi er sjølve begeistra over kor interessant e-NO burde kunne bli som forskingsverktøy, og vi vonar at norsk språkforskning på sikt vil utnytte moglegheitene som no opnar seg opp.

Når det gjeld brukargrensesnittet som vi må utvikle for e-NO, gler vi oss til å lage eit system som kjem nye og gamle brukarar i møte. Så langt har vi sett opp nokre ideal for dette brukargrensesnittet: Vi vil gjerne at søkjegrensesnittet skal kunne tilpassast ulike brukargrupper, slik at enkel bruk av e-NO skal kunne gjerast ved eit enkelt søkjeskjema, medan meir avansert bruk skal kunne gjerast ved å hjelp av skreddarsydde søkjeskjema. Vi ser førebels for oss to hovedtypar bruk som det må lagast tilpassa søkjesystem for: meir tradisjonell ordboksbruk der ein leitar etter individuelle oppslagsord og gjerne har moglegheit til å "bla" litt på måfå rundt i ordboka, dessutan meir systematisk ordboksbruk, der målet er å finne mønster på ulike nivå. Men sjølv om desse to typane bruk peikar seg ut, må individuell brukartilpassing vere ei rettesnor.

Det blir elles viktig for oss å leggje til rette for upresise eller "fuzzy" søk, slik at ein kan finne fram til ord ein ikkje alt kjenner. (Sjå De Schryver (2003:174) som viser til Abate 1986, Webster & Ning 1996, Nesi 2000 og Malmgren 1999:287ff.) Til hjelp her ser vi for oss trunkering, jøkerteikn og menyar med nesten-treff i samband med logisk utnytting av informasjonskategoriane i mikrostrukturen. Eit godt hjelpesystem blir viktig, og vi ser at vi truleg vil ha mykje å vinne på å få profesjonell hjelp til utforming av grensesnittet. Ikkje minst blir dette viktig for arbeidet med å presentere lenker til underlagsmaterialet for ordartiklane slik at ein lett kan bevege seg frå ordartikkel til setelarkiv og korpuskonkordansar og attende.

### 5.3 Mediostrukturen

Med mediostrukturen meiner vi her dei eksplisitte lenkene som går mellom etablerte einingar i databasen, både mellom mikro- og makrostruktureiningar. Vi har alt laga eit internt lenkesystem i databasen som primært vert utnytta til kryssreferansar. Kryssreferansane vert altså ikkje berre uttrykte i skrift, dei vert etablerte som fysiske koplingar mellom element i databasen. Førebels har vi utnytta dette lenkesystemet relativt tradisjonelt, dvs. vi har gjort koplingane med tanke på papiirutgåvebruken, og for å hjelpe redaktørane til å oppnå intern konsistens, dessutan for å hindre redundans og tomme peikarar. Men vi ser jo at i dette systemet ligg det moglegheiter for å lage meir gjennomtenkte nettverk på tvers av basen. På teoretisk grunnlag kunne systemet takast i bruk for å etablere onomasiologisk eller semantisk definerte nettverk, og sjølv om vi pr i dag ikkje har rom for å definere delprosjekt som går i denne leia, ligg det her moglegheiter for å etablere dotterprosjekt til *Norsk Ordbok*, t.d. ved å knyte doktorgradsprosjekt til redaksjonen.

## 6. Konklusjon

Vi er entusiastiske over dei moglegheitene som no opnar seg i og med realiseringa av ei elektronisk utgåve av *Norsk Ordbok*. Sjølv om formell avklaring må kome på plass, og vi enno har arbeid å gjere med omsyn til meir presis innpassing av tidlegare bokmanuskript i databaseformatet, ser vi at vi i grunnen har fått etablert ein databasestruktur og eit søkjesystem som det blir tilfredsstillande å gjere ålment tilgjengeleg. I denne fasen av arbeidet er vi opne for alle konstruktive innspel, og er slik sett takknemlege for å ha fått høve til å leggje fram preliminære planar og visjonar i eit nordisk leksikografifagleg forum.

## Referansar

- Abate, F.R. 1985: *Dictionaries Past & Future: Issues and Prospects*. I: *Dictionaries* 7, 270–283.
- Cederholm, Y. & L. Rogström 2001: Gamla ord blir som nya. OSA-databasen som källa och resurs. I: Gellerstam, M. et al. (red.): *Nordiska studier i lexikografi* 5. (Rapport från Konferens om lexikografi i Norden. Göteborg 27–29 maj 1999.) Göteborg: Nordiska föreningen för lexikografi, 66–79.

- De Schryver, G.-M. 2003: Lexicographers' Dreams in the Electronic-Dictionary Age. I: *International Journal of Lexicography* 16:2, 143–199.
- Gawinski, B. under trykking: Spansk mad – en illustred ordbog. Vil bli publisert i rapport frå den 8. konferansen om leksikografi i Norden, Sönderborg 24–28 maj 2005.
- Grønvik, O. 2005a: Redigeringsprogrammet for Norsk Ordbok 2014. I skjeringspunktet mellom menneskeleg skjønn og automatisering. I: R. V. Fjeld & D. Worren (red.): *Nordiske studiar i leksikografi* 7. (Rapport frå Konferanse om leksikografi i Norden. Volda 20.–24. mai 2003.) Oslo: Nordisk foreining for leksikografi, 156–165.
- Grønvik, O. 2005b: Norsk Ordbok 2014 from Manuscript to Database – Standard Gains and Growing Pains. I: F. Kiefer, G. Kiss & J. Pajzs (red.): *Papers in Computational Lexicography. Complex 2005*. Budapest: Linguistic Institute, Hungarian Academy of Sciences, 60–70.
- Kruijsen, J. & J. Swanenberg 2004: Les dictionnaires du brabançon (WBD) et du Limbourgeois (WLD) sur l'internet. I: G. Williams & S. Vessier (red.): *Eleventh Euralex international congress. Le Paquebot/Lorient/France. Proceedings*. Lorient: Université de Bretagne-Sud, 117–126.
- Løfberg, L. et al. 2004: Using a semantic tagger as a dictionary search tool. I: G. Williams & S. Vessier (red.): *Eleventh Euralex international congress. Le Paquebot/Lorient/France. Proceedings*. Lorient: Université de Bretagne-Sud, 127–134.
- Malmgren, S.-G. 1999: Att göra en enspråkig definitionsordbok mer produktionsinriktad – exemplet Nationalencyklopedins ordbok. I: Slotte, P. et al. (red.): *Nordiska studier i lexikografi* 4. (Rapport från Konferensen om lexikografi i Norden. Esbo 21–24 maj 1997.) Helsingfors: Nordiska föreningen för lexikografi, 285–290.
- Nesi, H. 2000: *The Use and Abuse of EFL Dictionaries. How learners of English as a foreign language read and interpret dictionary entries*. (Lexicographica. Series Maior 98.) Tübingen: Max Niemeyer Verlag.
- Pelissier, C. & C. Becker 2005: Adaption of a Computerized Dictionary for Language Learning: The “*Trésor de la Langue Française informatisé*” and French Language. I: F. Kiefer, G. Kiss & J. Pajzs (red.): *Papers in Computational Lexicography. Complex 2005*. Budapest: Linguistic Institute, Hungarian Academy of Sciences, 161–170.
- Ruus, H. 2002: A Corpus-based Electronic Dictionary for (Re)search. I: A. Braasch & C. Povlsen (red.): *The Tenth Euralex International*

*Congress. Copenhagen – Denmark. August 13–17, 2002. Proceedings.* København: Center for Sprogteknologi, 175–185.

Kristin Bakken  
prosjektdirektør/dr.art.  
Norsk Ordbok 2014  
P.boks 1021 Blindern  
N-0315 Oslo  
[kristin.bakken@iln.uio.no](mailto:kristin.bakken@iln.uio.no)

Christian-Emil Ore  
fagleg leiar/dr.art.  
Eining for digital dokumentasjon  
P.boks 1123 Blindern  
N-0317 Oslo  
[c.e.s.ore@edd.uio.no](mailto:c.e.s.ore@edd.uio.no)