

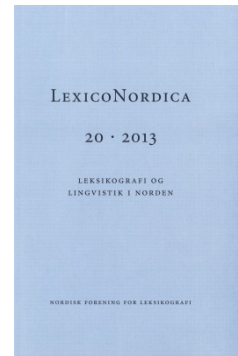
# LexicoNordica

Forfatter: Øyvind Eide [*Solid utgivelse om elektronisk leksikografi*]

Anmeldt værk: Sylviane Granger & Magali Paquot (red.): *Electronic Lexicography*.  
Oxford: Oxford University Press 2012. 517 sider. Pris: 75 GBP.

Kilde: LexicoNordica 20, 2013, s. 205-222

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.

# Solid utgivelse om elektronisk leksikografi

Øyvind Eide

Sylviane Granger & Magali Paquot (red.): *Electronic Lexicography*.  
Oxford: Oxford University Press 2012. 517 sider. Pris: 75 GBP.

## 1. Innledning

*Electronic Lexicography* er et resultat av et forskningsprosjekt finansiert av det belgiske *Fonds de la Recherche Scientifique* (FNRS). Den er plassert i grenseland mellom leksikografi og digital humaniora. Denne anmelderen har sin bakgrunn fra det sistnevnte området, noe anmeldelsen nok vil bære preg av. Boka består av 20 kapitler skrevet av ulike forfattere. Etter innledningen er de øvrige kapitlene delt inn i tre deler: først en generell del om elektronisk leksikografi, så en del som beskriver et antall prosjekter, og til slutt en del som fokuserer særlig på forholdet til brukerne.

Dette er en solid og god utgivelse som inneholder et vell av nyttig informasjon. Den kan leses med stort utbytte fra perm til perm, men også enkeltdele eller enkeltkapitler kan leses separat. Jeg vil først gi en veiviser til boka, der jeg går gjennom tematikken i de ulike kapitlene. Jeg vil så komme med noen kritiske bemerkninger til boka som helhet. Jeg kommer derimot ikke til å gå inn i de mange diskusjonene de ulike kapitlene åpner for, til det er boka for vidtfavnende.

## 2. De enkelte kapitlene

Grangers innledning framstiller historien til den elektroniske leksikografien og definerer noen grunnbegreper, før seks viktige nyvinninger blir beskrevet: integrasjon av korpus, større og bedre datatilfang, tilgangseffektivitet, brukertilpasning, hybridisering og bidrag fra ordboksbrukerne. Disse seks punktene utgjør også et fokus for boka som helhet. Et annet fokus for hele boka er det innledningen avsluttes med: Vi står nå ved et vendepunkt i leksikografiens historie, vi er i en tid hvor tekniske og økonomiske forhold sammen peker mot store endringer i hvordan ordbøker skrives og formidles.

### 2.1. Leksikografien ved et vannskille

Første del har nettopp dette som fokus: leksikografien ved et vannskille. Rundells kapittel tar utgangspunkt i redaktørens rolle og gir en klargjørende oversikt over de mulighetene som finnes. Han ser for seg at maskinene vil ta over flere og flere av de leksikografiske oppgavene og at redaktørens rolle snarere blir å godkjenne valg og utvalg gjort av programvaren enn å skrive ordbøker. Det er klart for Rundell at ordbøker har funnet et ideelt medium i form av nettpublikasjoner, men hvordan skaffe inntekter? Han ser utviklingen som overveiende positiv og frykter ikke at ordbøkene vil forsvinne, men ser for seg en overgang til ordbøker som tjenester snarere enn produkter i framtida, en overgang vi allerede ser starten av.

Bidraget fra Kilgarriff & Kossem fokuserer på korpusverktøy, med hovedvekt på *Sketch Engine*. Kapitlet gir først og fremst en oversikt over verktøyet, de ulike funksjonene, og hvordan det er bygd opp. Basert på dette setter de opp noen kriterier for hva leksikografer trenger fra sine verktøy. Veksten i korpusstørrelse gjør slike verktøy viktigere og viktigere. Men de tror ikke at automatiske

korpusverktøy vil nå et nivå der de erstatter ordbøker. De skriver også om integrasjon og tilpasning av verktøy. Konklusjonen er at korpus ikke vil erstatte ordbøker, de gir nyttige sammendrag. Men når ordboka kommer til kort, kan brukeren også ha nytte av å ha tilgang til korpus. Kapitlet gir en grei gjennomgang av verktøyene som beskrives, men er litt for snevert til å fungere godt som et generelt innledningskapittel.

Hanks legger i sitt kapittel vekt på den tette integrasjonen mellom elektroniske ordbøker og deres kilder, konkret mellom ords enkeltbetydninger og eksempler på faktisk språkbruk. Korpus kan fungere som en løsning på et tradisjonelt fravær av bruksbasert dokumentasjon. Han hevder at pre-korpusbasert leksikografi «was little more than a collection of stabs in the dark» (s. 63), og kritiserer en tradisjonell overfokusering på etymologi. Store korpus gir enspråklige ordbøker nok evidens til å avgjøre hva som skal tas med og hva som skal utelates, og korpusbaserte ordbøker tenderer til å konvergere i hva de sier om språk, også når de er basert på ulike korpus. Han diskuterer hvordan dette griper inn i teoretiske diskusjoner om mening og konkluderer med at all seriøs leksikografi framover vil være korpusdrevet og ikke basert på sitatsamlinger eller gjetninger og spekulasjoner. Argumentasjonen hans er skarp og poengtert, og jeg synes han underbygger konklusjonen godt.

Abel gir en oversikt over ordbokssystemer, dvs. systemer som brukes til å skrive ordbøker, inkludert tilleggsverktøy og -funksjoner som kan kobles til «rene» ordbokapplikasjoner. Hun skildrer generelle trender, noe som dekker de fleste systemene som finnes fordi det er en tydelig konvergens, særlig mellom kommersielle systemer. De fritar leksikografen fra jobben med å passe på stildetaljer; et viktig mål med programvaren har vært og er å gjøre det lettere å skrive artikler. I tillegg kommer sikker lagring av data og effektiv prosjekthåndtering. Hun går gjennom definisjoner av hva et slikt system skal være og hva det bør inneholde. Hun disku-

terer forholdet mellom selvutviklede og ferdigkjøpte systemer og sammenfatter sakskomplekset på en god måte. Imidlertid synes jeg hun overfokuserer på forskjellene mellom de to alternativene, særlig for større ordbøker. Hun sier riktignok at IT-ekspertise er nødvendig også når man kjøper systemer, men det er viktig å gjøre det klart at det ikke bare er snakk om driftspersonale. Man må gjøre endringer av hylleware for å tilpasse den til store, komplekse prosjekter, enten det skjer hos produsent eller bruker. På den andre siden er såkalt selvutviklede systemer i dag også basert på sammenkoblede standardmoduler, så som databaser og grafiske verktøy. Begge prosessene er arbeidskrevende og kompliserte.

Tarp skriver i sitt kapittel om overgangen fra trykte ordboksverk til elektroniske ordboksverktøy. Trenger man ny teori eller kan man bruke teori etablert i det tryktes tidsalder? Selve ideen om leksikografisk teori er omdiskutert, men Tarp tar utgangspunkt i at teorier eksisterer og er nødvendige. Spørsmålet er om man kan justere teoriverket ut fra overgangen til det digitale eller om det er snakk om et paradigmeskifte. Han ser på leksikografi som en type informasjonsvitenskap og ser en konvergens der ulike referanseverktøy smelter sammen og der også den leksikografiske funksjonsteorien kan finne sin plass. Dette er en videreutviklet funksjonsteori: Det er ikke typer av brukere som konsulterer typer av ordbøker, det er snakk om konkrete (partikulære) hendelser. Individualisering og brukertilfredshet må settes i fokus. Dette kan løses ved avanserte digitale verktøy. Han avslutter med en klassifisering av Internett-baserte verktøy der de mer avanserte er hypotetiske systemer som ennå ikke finnes, som ordboksartikler med dynamiske data og individuelle løsninger basert på dette. Slike løsninger må utvikles i tverrfaglig samarbeid mellom informatikere, informasjonsspesialister og leksikografer, og må være basert på en avansert teoretisk forståelse.

Prinsloos artikkel, som avslutter første del, representerer et interessant valg av redaktørene i den forstand at den tilsynelatende

presenterer seg som en lokal artikkel som dreier seg om leksikografi i Sør-Afrika. Imidlertid viser dens plassering implisitt til at de øvrige artiklene i bokas første del også i stor grad er lokale, men de er lokale i det som i dag er dominante deler av verden.<sup>1</sup> For artikkelen er ikke en lokal diskusjon av sørafrikanske forhold alene, det er en framstilling av hvor man står som leksikograf i en virkelighet der ulike språk er svært ulike hva ressurser angår. Den gir en oversikt over situasjonen i Sør-Afrika, men dette gjøres som et eksempel på de store delene av verden der også ressursfattige språk brukes. Når dette er formålet, kunne man nok ønsket noe flere koblinger til leksikografisk arbeid utenfor Sør-Afrika, særlig for språk som snakkes i flere land eller er nært beslektet med språk i andre land. Prinsloo spør om elektroniske ordbøker kan åpne muligheter som det trykte mediet ikke gir. For bantuspråk ligger det klare muligheter til bedre løsninger på noen av problemene man har hatt i trykte ordbøker, ikke minst identifikasjon av ordstammer og problemene med gjenfinning dette fører til. Men konklusjonen er at elektroniske ordbøker i stor grad fortsatt er utformet som digitale versjoner av trykte formater, og derfor ikke tar ut det potensialet som finnes.

Første del som helhet gir en innføring i noen grunnproblemer som man kommer tilbake til senere i boka. Som en slik innføring fungerer den jevnt over bra, selv om ikke alle kapitlene har like stor bredde. Særlig Kilgarriff & Kosem kunne nok passet like godt i neste del, som en prosjektbeskrivelse.

## 2.2. Prosjekter

Denne delen utgjøres av en serie kapitler som hvert går gjennom ett eller flere nyskapende prosjekter. Det første er skrevet av Ver-

---

1 Det angloamerikanske fokuset man ofte finner i engelskspråklige utgivelser, er derimot fraværende i denne boka, noe man må regne med skyldes bokas opphav i et belgiskfinansiert prosjekt.

linde & Peeters og handler om *Interactive Language Toolbox* (ILT), som er utviklet basert på tilbakemeldinger fra brukere av tidligere versjoner av systemet *Base lexicale du français* (BLF). Målet er å samle nok data om ordboksbrukernes behov til å gjøre fornuftige utvalg av tilgjengelige leksikografiske ressurser for de ulike brukerne, slik at de kan møte enkle, tilpassede grensesnitt. Kapitlet preges noe av salgsspråk, men ikke mer enn at det fungerer som en god gjennomgang av systemet.

Paquots kapittel beskriver *Louvain English for Academic Purposes Dictionary* (LEAD), et integrert ordboks- og korpusverktøy for fremmedspråksopplæring og -skrivestøtte. Et hovedmål var å lage et system der brukerne fikk tilgang til så gode og relevante eksempler som mulig. Dette mener de å ha løst ved å lage systemet genrebasert og fagspesifikt, samtidig som de aktivt bruker analyse av data fra språklæringskorpus. Spørsmålet er om brukerne er tilstrekkelig interessert i disse avanserte løsningene til at de kan konkurrere med enklere systemer som er lett tilgjengelige på Internett.

Kübler & Pecman skriver om *ARTES*, ei fransk-engelsk ordbok for fagspråk. Som en konsekvens av engelsk som lingua franca er det et voksende behov for tospråklige hjelpemidler. *ARTES* er bygd for studenter og kombinerer ei tospråklig fagordbok med en database over termer og fraser.

Neste kapittel, skrevet av L'Homme, Robichaud & Leroyer, er om *DiCoInfo*, en nettbasert termressurs for informatikk og Internett. Hovedtyngden er fransk, men også engelsk og spansk er på vei inn. Opprinnelig var dette et forskningsverktøy, men en utvidet versjon er senere utviklet for studenter og oversettere, der termbaserte kollokasjoner står i fokus. Forfatterne beskriver en brukerundersøkelse de har gjennomført, som viser at de uten å endre strukturen i den opprinnelige forskningsdatabasen kunne utvikle brukervennlige verktøy på toppen.

Tasovac framstiller i sitt kapittel ei ordnettbasert ordbok som en løsning på flere leksikografiske problemer, inkludert hvordan

man finner ord man ikke kjenner til på forhånd, gjennom å overkomme alfabetiseringens begrensninger. Han beskriver ei serbisk ordbok knyttet til annoterte digitale utgaver av skjønnlitteratur, *Transpoetika*, men målet er videre enn som så; han ønsker å kunne bruke ordnett til å bygge bro over gapet mellom tradisjonell leksikografi og datalingvistikk (Natural Language Processing, NLP). Han mener også at lisensbestemmelsene for ordnett er nyttig for prosjekter uten stabil finansiering, særlig i forbindelse med ressursfattige språk som serbisk. Dette styrkes ved at *Transpoetika* går ut over begrensningene i en ordnettbasert plattform ved å knytte til webtjenester som kan brukes av andre websider og brukermedvirkning (crowdsourcing), i tillegg til eksterne gratissystemer for kart, bilder og lyd. For framtida foreslår han at flere systemer kan integreres inn i denne rammen.

Neste kapittel, skrevet av Meyer & Gurevych, tar opp den nettbaserte ordboka *Wiktionary*, som er ordbøkernes versjon av Wikipedia, skrevet på dugnadsbasis. Denne omtales kritisk i et tidligere kapittel (Hanks), men dette kapitlet gir et mer positivt bilde, delvis basert på kvantitative sammenligninger med andre elektroniske ordbøker for engelsk, tysk og russisk. Man kan stille spørsmål ved metodene de har brukt, men en velfundert kvalitativ kritikk av en så omfattende tjeneste som Wiktionary ville vært en omfattende oppgave. Det mest overraskende resultatet av sammenligningene er hvor lite overlapp det er mellom Wiktionary og de andre ordbøkene, både når det gjelder oppslagsord og betydninger. Dette nærmest roper etter videre forskning, samtidig som det kan være lurt av ordboksredaktører å vurdere sin egen lemmaseleksjon og sine egne betydningsbolker opp mot Wiktionary. En styrke ved Wiktionary over de fleste andre ordbøker er at alle artikler har redigeringshistorikken åpent tilgjengelig, noe som er nyttig for å undersøke hvordan artikler utvikles. Et annet godt poeng er at Wiktionary kan brukes som et enkelt verktøy for lingvister som studerer truede språk og ønsker å publisere forskningen sin. Generelt ser



Meyer & Gurevych på Wiktionary og «collaboratory lexicography» som et nyttig supplement til tradisjonell leksikografi.

Siste kapittel i del to er skrevet av Kristoffersen & Troelsgård og omhandler elektronisk leksikografi for tegnspråk. De bruker *Ordbog over Dansk Tegnsprog* som sitt hovedeksempel. De gir først en innføring i tegnspråk og tegnspråksleksikografi, som er en ung disiplin. Hovedproblemet er hvordan man kan framstille tegn når man ikke har et skriftspråk å gjøre det i. De går gjennom ulike muligheter som alle har sine fordeler og ulemper: video, animasjoner, bilder, tekstlige representasjoner og formelle notasjonssystemer. Framstilling av tegnene er en ting, men hvis man skal ha ei reelt enspråklig ordbok må den også være søkbar, og dette er et vanskeligere problem å løse. Elektroniske tegnspråksordbøker er for det meste søkbare gjennom skrevne ord, selv om enkelte også tilbyr søk på fonologiske kriterier så som hendenes stilling. De viser så hvordan de har forsøkt å finne løsninger på disse problemene i *Ordbog over Dansk Tegnsprog*, fulgt av noen vyer for framtida. Et interessant spørsmål er om og i tilfelle på hvilken måte man kan lage funksjonelle korpus for tegnspråk.

Del to er ment å omhandle et antall enkeltprosjekter, og hvert kapittel har derigjennom et snevrere fokus. Likevel synes jeg i det store og det hele at det er en interessant samling av kapitler som i større eller mindre grad peker ut over de prosjektene som beskrives.

### 2.3. Ordboksbrukeren

I denne delen av boka er brukerperspektivet i fokus. Dette har vært godt dekket i de to første delene også, men her er det hovedtema for kapitlene. Først ute er Dziemianko med en sammenligning mellom papirordbøker og elektroniske ordbøker som to ulike medier for presentasjon av språkmateriale. Hun kommer inn på grunnleggende forskjeller mellom de to mediene i seg selv, så som oversiktighet og hvordan kryssreferanser fungerer, men tar også

opp forskjeller i hvordan de to mediene oppfattes, for eksempel når det gjelder troverdighet. Hun gjør en metastudie av forskning publisert i 2000–2011. Resultatene hun går gjennom, viser stort sprik på mange områder, og hun konkluderer med at det er behov for forskning som kan gi oss et klarere bilde av forskjellene mellom ulike typer ordboksbrukere og mellom ulike typer bruk av ordbøker. Hun foreslår også bedre komparative undersøkelser av hvor nyttige ulike typer av elektroniske ordbøker er. Det er et åpent spørsmål om man kan finne noe generelt svar på om elektroniske ordbøker eller papirordbøker er best.

Lew spør hvordan elektroniske ordbøker kan gjøres mer effektive, men ser et fravær av eksperimentell forskning som grunnlag for hvordan man best kan lage dem. Han mener elektroniske ordbøker først og fremst kan overgå trykte i hvor effektivt de gir brukeren tilgang til leksikografiske data, og kanskje også i bruk av multimedia. Det første er hovedtema i kapitlet, og dreier seg blant annet om i hvilken grad man ledes til rett ord når man feilstaver. Dette er jo et problem som aldri kan løses fullt ut, men der mye kan gjøres. Lew går gjennom ei rekke ulike måter å løse tilgangsproblemet på i elektroniske så vel som papirbaserte ordbøker, og oppsummerer med at mye kan gjøres med små investeringer og dagens teknologi.

Neste kapittel, skrevet av Nesi, tar opp alternative elektroniske ordbøker, eller AED. Disse ressursene, som for leksikografer kan framstå som tilfeldige samlinger av kilder med uklart proveniens, har lav status og er forsket lite på, men de har mange brukere. Kapitlet er et forsøk på å ta dem på alvor og synliggjøre problemer som oppstår når brukere, særlig elever og studenter, stoler for mye på slike produkter. Artikkelen fokuserer på asiatiske produkter for engelsk. Jeg er usikker på om tilsvarende problemer finnes for fremmedspråkoplæring i Norden, men det er nok et område det er viktig å holde øye med.

Bowkers kapittel tar for seg elektroniske ordbøker som et

hjelpemiddel for oversettere, og spør om det tradisjonelle kjær-  
 lighet-hat-forholdet oversettere har hatt til ordbøker, kan endres  
 ved at digitale ressurser møter oversetternes behov bedre. Hun  
 går gjennom noen mulige løsninger, hvorav en del er under ut-  
 vikling, mens andre foreløpig kun er på ideplanet. En hovedvekt  
 ligger på forholdet til terminologi, særlig fagspråk, som er viktig  
 å raskt kunne sette seg inn i for oversettere. I hvilken grad skal  
 generelle ordbøker inkludere fagspesifikke termer? Er det andre  
 relevante forskjeller mellom ulike fagspråk, så som ulike pragma-  
 tiske strategier, som ordbøker kan bidra til å forstå bedre? Halvveis  
 bearbeidet korpusinformasjon, som ordskisser,<sup>2</sup> er ett eksempel på  
 materiale hun tror oversettere kunne ha nytte av bedre tilgang til.  
 Hun råder leksikografer til å følge med på utviklingen i oversetter-  
 studier for å kunne lage bedre verktøy for oversettere.

Fuertes-Oliveras kapittel tar opp igjen tematikken fra flere tid-  
 ligere kapitler og ser på gratisordbøker<sup>3</sup> på Internett som verktøy  
 for undervisning i forretningsengelsk. Han tar også opp en annen  
 tråd som har vært diskutert i tidligere kapitler, nemlig å se elektro-  
 niske ordbøker som en type informasjonssystemer og bruke teori  
 fra det fagområdet, så som prinsippet om brukervennlighet (usa-  
 bility) for dataverktøy. Basert på Nesis kapittel setter han opp seks  
 kriterier for brukervennlighet knyttet til gratisordbøker på Inter-  
 nett. Han konkluderer med at problemene med ressursene i stor  
 grad må løses gjennom bedre opplæring av brukerne. De må få en  
 bedre forståelse av forskjellen mellom autorative og mindre au-  
 torative kilder og bruke ressursene med vett og forstand. For selv  
 om gratisordbøkene ikke er i front når det gjelder leksikografisk  
 kvalitet, er de likevel nyttige for brukere i gitte situasjoner.

Siste kapitlet i boka er skrevet av Müller-Spitzer, Kopenig &

2 *Sketch Engine*, se Kilgarriff & Kosems kapittel omtalt ovenfor.

3 Det engelske ordet *free*, som brukes i Fuertes-Oliveras overskrift, er ambivalent og kan bety både 'fri' og 'gratis'. Forfatteren påpeker også at *free* i denne sammenhengen er vanskelig å definere. I dette kapitlet ser det ut til at hovedfokus er på betydningen 'gratis'.

Topel og går gjennom funn fra et empirisk forskningsprosjekt på bruk av nettordbøker. Gjennom fire punkter fra brukerundersøkelser klargjør forfatterne hvilke ønsker og krav brukerne har. Et interessant funn er at et flertall i alle brukergruppene prioriterer pålitelighet høyest. Metoden de bruker, har sine svakheter, blant annet fordi deltakerne i slike undersøkelser ikke vet hva som er mulig hvis de ikke har sett det før, noe som delvis kan forklare at nyskapende presentasjonsformer ble ansett som mindre viktig. Likevel argumenterer forfatterne godt for at slike undersøkelser er nyttige ved siden av andre metoder, som analyser av loggfiler. Den grundige metodikken de har lagt til grunn bør kunne gjøre dette til et nyttig kapittel for utviklere av nettordbøker.

Del tre som helhet gir en bred dekning av brukerperspektivet ved å trekke inn ulike brukergrupper og ulike typer verktøy. Det jeg savner her, er en oppsummering. Det kunne gjerne vært en kort oppsummering av hver av de tre delene, men i hvert fall burde boka hatt et helhetlig oppsummeringskapittel som trekker trådene fra innledningen gjennom de ulike kapitlene. Slik det nå står, henger det litt i løse lufta for de leserne som leser boka fra perm til perm. For lesere av enkeltkapitler, som det kan bli mange av, er dette selvsagt ingen viktig innvending.

### 3. Generelle tendenser

Dersom man ønsker å prøve ut nye løsninger og ønsker gode ideer, er det mye å hente i *Electronic Lexicography*, samtidig som flere av kapitlene har grundige gjennomganger av tidligere forskning. Det er i det hele tatt ei bok med mange sterke sider.

Brukerfokus står sentralt, og boka kan fungere godt som et referanseverk for brukertilpasning. Brukertilpasning vil i denne sammenhengen kunne bety minst to ting. Det er for det første snakk om tilrettelegging og spesialiserte tjenester for ulike grup-

per av brukere. Men det kan også dreie seg om bidrag fra brukerne som samles inn gjennom dugnadssystemer (crowdsourcing). Slike dugnadsopplegg, som kan være åpne for spesielle grupper så vel som for allmennheten, er tatt med i flere av systemene som beskrives. Her er mye under utprøving, og det som beskrives, er i like stor grad skisser og ideer som det er ferdige systemer. Visjonene for hva som kan komme i framtida, er mange og avanserte. Man kan ane en utålmodighet hos flere av forfatterne over at utviklingen i retning av brukertilpasning og interaktivitet i elektroniske ordbøker ikke går raskere.

Mange av vyene for digital produksjon av ordbøker er spennende, men jeg tror det vil utvikle seg et skille mellom større og mindre språk, noe for eksempel Prinsloo antyder. Dette er relevant også i en europeisk kontekst. Det er ikke nødvendigvis slik at selv de store nordiske språkene er store nok til å rettferdiggjøre den innsatsen som må til for å få til de avanserte løsningene som skisseres i de ulike kapitlene. For språk som samisk og kvensk virker det iallfall usannsynlig.

Det er fra et nordisk perspektiv også interessant å se hvordan de kapitlene som beskriver rasjonalisering av leksikografisk arbeid gjennom maskinelle rutiner, ikke i det hele tatt nevner de konsekvenser dette vil ha for de ansatte. Heller ikke tas slike perspektiver med når overgangen fra fast ansatte til hjemmearbeidende frilansere beskrives. Dette er vel et uttrykk for at forholdene i arbeidslivet er tøffe utenfor Norden, ikke minst i våre dager.

Mange av kapitlene omhandler enkeltprosjekter og enkeltordbøker, og de er gjerne skrevet av redaktører eller andre medarbeidere. Det kan nok føre til noe mindre kritisk distanse enn det en eksternt ville hatt, men det er et forståelig valg; det gjør at forfatterne har god innsikt i det de skriver om. I ei bok som dette blir dette balansert av at det er mange kapitler med ulike forfattere. Likevel er dette viktig å ha i bakhodet, særlig om man leser enkeltkapitler fra boka.

Det ligger implisitt i flere av kapitlene og eksplisitt hos Abel

at XML er et grunnlag for ordboksredigering; hos sistnevnte er dette uttrykt ved at datamodeller generelt likestilles med XMLs formelle dokumentstruktursystem DTD. Hennes gjennomgang er i det hele tatt tett knyttet til XML som grunnstruktur, noe som er i tråd med de fleste systemer man ser i dag: XML representerer jo en kontekstfri grammatikk<sup>4</sup> som egner seg godt for den trestrukturen man gjerne ser i artikler.

En ting er at dette er litt gammelmodig og at man heller burde brukt nyere XML-formalismer som Schema eller RNG. Viktigere er det at dette sier noe om hva man ser ei ordbok som. Det er helt klart at ei ordbok kan sees på som strukturert som et tre. Det betyr at en trestruktur er et godt hjelpemiddel i bruk og produksjon av ordbøker. Men fordi trestrukturer først og fremst egner seg for selve artikkelen, blir dette et uttrykk for et syn på ordbokssystemer som rene artikkelsystemer frikoblet fra sin kontekst. Dette er noe som flere av kapitlene i boka, Abels inkludert, advarer mot.

Det som mangler, er en grunnleggende diskusjon om prinsippene bak de datastrukturene man velger i elektronisk leksikografi. Boka som helhet er glimrende i det den er god i, nemlig som en praktisk veiledning. Den inneholder et vell av fakta som for den leseren framstår troverdig, gitt de redaksjonelle valgene som er gjort, og den tar opp teoretiske problemstillinger innen leksikografi, ikke minst i Tarps kapittel. Men selve spenningspunktet mellom leksikografi og det digitale, mellom «electronic» og «lexicography», blir mer beskrivende enn teoretiserende. Boka går ikke i tilstrekkelig grad inn i grunnleggende problemer med ordboksformater knyttet til de datastrukturene man velger. Er trestrukturer de beste for prosjektstyringsdelen av systemet, eller egner en relasjonsdatabasestruktur seg bedre for å holde kontroll med brukere og deres rettigheter, produksjon, og så videre? Og hva med kilde-materialet? I digital edisjonsfilologi har det vært en lang diskusjon

---

4 Om kontekstfrie og andre typer grammatikker, se Grune & Jacobs (2008).

om problemene med XML-baserte formalismer.<sup>5</sup> Har dette konsekvenser for kildematerialet for ordbøker, korpus så vel som andre typer kilder? Det er også et spørsmål om trestrukturer er den beste modus for brukergrensesnittet. Müller-Spitzer, Koplenig & Topels artikkel viser at både ei rekke med tab-paneler og all informasjon på ett panel vurderes som bedre av brukere av nettordbøker enn en trestruktur som tilsvarende utforskeren i Windows.

Alle er vel i dag enige om at det lineære formatet til en tradisjonell trykt tekst er problematisk når det gjelder å uttrykke kunnskap om språk. Man ønsker seg andre typer strukturer, og de av oss som har arbeidet med å analysere ordbøker for overgang til elektronisk form, har tydelig sett en implisitt trestruktur i før-elektroniske trykte ordboksartikler. Men ved å løse det strukturelle linearitetsproblemet i den trykte teksten ved å beholde lineariteten i tekstbrokkene (man skriver fortsatt «definisjon» med bokstavene i riktig rekkefølge) og legge til en eksplisitt trestruktur, har man gjort seg til offer for en ny tvangstrøye. Denne tvangstrøyen er romsligere enn den gamle, men den er fortsatt der. Det tydeligste eksemplet er kryssreferansene, som har lett for å ende opp som noe som kommer i tillegg til den «ekte» strukturen. Nødvendig, men ikke tilstrekkelig, er Tarps påpekning av forskjellen mellom database og ordbøker som kan genereres basert på databasen.

En av «ordbøkene» som omtales er ordnett, mest hos Tasovac, men også i andre kapitler. Det er interessant å se hvordan ordnett framstår fra et leksikografisk utgangspunkt, og Tasovacs forsøk på en brobygging mellom ordnett og tradisjonelle ordbøker er prisverdig. Jeg skulle likevel ønske at man hadde med et kapittel som

5 Renear et al. (1996) argumenterer for at tekster generelt er strukturert på en måte som gjør det naturlig å kode dem inn i kontekstfrie grammatikker. Motstridende syn framholdes av blant andre McGann (2001). Konsensus i digital edisjonsfilologi i dag er at trestrukturer representerer et problem, men det er mulig å finne pragmatiske løsninger innenfor rammen av XML (TEI Consortium 2012, kap. 20).

tok opp den grunnleggende strukturelle forskjellen mellom ordnetts grafstruktur og trebaserte ordbøker.

Et annet problem i tråd med dette tar Kristoffersen & Troelsgård opp i sitt kapittel om tegnspråk. Hvordan kan man organisere ei ordbok over et språk som ikke har et skriftspråk? Dette løses jo for muntlige språk ved å lage et skriftspråk, lage ordbøkene i skriftspråket, og så komplettere resultatet med f.eks. transkribert tale og lydopptak. Dette er en løsning som også ofte brukes for tegnspråk, men siden tegnspråk er et annet språk enn skriftspråket, blir ordboka da uvegerlig ei tospråklig ordbok. Kristoffersen & Troelsgård har en del forslag til praktiske løsninger, men jeg savner en tydeligere diskusjon om forholdet mellom praktiske løsninger og prinsipielle muligheter. Er det i det hele tatt mulig å lage ei enspråklig ordbok over et språk som er så entydig knyttet til tid i sin modalitet?<sup>6</sup> Er dette et grunnleggende uløselig medieproblem, eller er det et problem som kan løses gjennom teknologisk utvikling? Kristoffersen & Troelsgård påpeker at den første enspråklige tegnspråksordboka der tegnspråk også er metaspråk for ordboka, ennå ikke er laget. Kan det være at den ikke er laget fordi det ikke er mulig?

Jeg undres også litt over at man kaller fagfeltet «electronic lexicography». Begrepet «digital» er både diskutabelt og omdiskutert, men det er mye brukt i søsterdisipliner som «digital history» og «digital arts and culture», i tillegg til sentrale felt som «digital media» og «digital humanities». Dette er kanskje annerledes på engelsk, men «elektronisk ordbok» gir meg assosiasjoner til en fagordbok for elektronikk. Det er kanskje en svakhet ved denne anmelderens begrepsforståelse, men gitt temaet for boka hadde en diskusjon av navnet på fagområdet vært på sin plass.

En teknisk svakhet er at mange av figurene som viser utsnitt av skjermbilder, er gjengitt i for liten størrelse og dels i svak kontrast, noe som fører til at de ikke kan leses godt nok til egentlig å støtte

6 For en oversikt over modaliteter i medieuttrykk, se Elleström (2010).



opp under hovedteksten. Dette er særlig et problem i Paquots kapittel, der nærmest uleselig tekst på figurer er en del av argumentasjonen.

#### 4. Oppsummering

Forrige avsnitt fokuserte på mangler ved boka. Noen slike finnes, men jeg ønsker ikke å gi inntrykk av at feil og mangler er det som preger denne utgivelsen. Både innholdsmessig, språklig og teknisk er dette en solid bokproduksjon. Selv om ulike forfattere har ulik språkføring, er det generelle språklige nivået høyt, og boka har et enhetlig preg. Boka er svært rik og inneholder mye mer som bør diskuteres videre enn de få aspektene jeg har funnet plass til å ta opp her. Referansene mellom kapitlene er mange, kun i enkelttilfeller har jeg observert at det glipper og det finnes henvisninger ut av boka på steder der det også burde vært henvisninger til andre kapitler.

Boka er god på det den er god på. Jeg synes redaktørene har vært svært dyktige i å framskaffe en enhetlig tekst som er svært leseverdig også for de av oss som ikke har spesialkompetanse på mange av områdene som omhandles. Jeg skulle ønske meg en tydeligere teoretisk diskusjon om grunnleggende strukturer bak ordbøker, gjerne i et eget kapittel. XML-problematikken nevnt ovenfor kan tjene som ett eksempel på hva som kunne vært tatt opp i et slikt kapittel. Jeg savner også et avsluttende kapittel som samler alle de spennende trådene som er spunnet ut.

## Litteratur

- Elleström, Lars (2010): *The Modalities of Media: A Model for Understanding Intermedial Relations*. I: Lars Elleström (red.): *Media Borders, Multimodality and Intermediality*. Basingstoke: Palgrave MacMillan, 11–48.
- Grune, Dick & Cerial J.H. Jacobs (2008): *Parsing Techniques: A Practical Guide*. 2. utg. New York: Springer.
- McGann, Jerome J. (2001): *Radiant Textuality: Literature after the World Wide Web*. New York: Palgrave.
- Renear, Allen, Elli Myonas & David Durand (1996): *Refining our Notion of What Text Really Is: The Problem of Overlapping Hierarchies*. I: Susan M. Hockey & Nancy M. Ide (red.): *Selected Papers from the ALLC/ACH Conference, Christ Church, Oxford, April 1992*. Oxford: Clarendon Press, 263–280.
- TEI Consortium (2012): *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. [2.3.0]. [17. januar 2012]. TEI Consortium. <<http://www.tei-c.org/Guidelines/P5/>> (juni 2013).

Øyvind Eide  
overingeniør, dr. phil.  
Universitetet i Oslo  
Enhet for digital dokumentasjon  
Institutt for lingvistiske og nordiske studier  
Postboks 1123 Blindern  
NO-0317 Oslo  
oyvind.eide@iln.uio.no