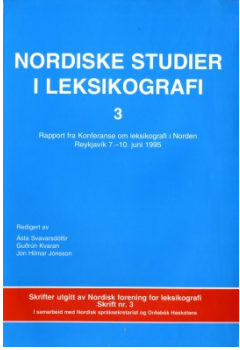


NORDISKE STUDIER I LEKSIKOGRAFI

| | | |
|------------|---|---|
| Titel: | Multimedialeksikografi: nye udfordringer for leksikografisk teori |  |
| Forfatter: | Pål Steigan | |
| Kilde: | Nordiske Studier i Leksikografi 3, 1995, s. 375-382 Rapport fra Konferanse om leksikografi i Norden, Reykjavík 7.-10. juni 1995 | |
| URL: | http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive | |

© Nordisk forening for leksikografi

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 Nordiske studier i leksikografi (1-5) 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.

Multimedialeksikografi: nye utfordringer for leksikografisk teori

Når vi bruker ordet bok, har vi alle en noenlunde klar forestilling om hva vi mener. Det er en gjenstand som består av trykte sider, bundet sammen i ryggen og gjerne med et omslag eller innbinding. Vi sier at en forfatter har skrevet ei bok, og mener da at han har skrevet en tekst som blir presentert i trykk mellom to permer. Vi er så vant med det halvtusenårgamle ekteskapet mellom boktrykkerbransjen og ordets folk at vi knapt tenker over at våre forestillinger er farget av teknologien. Ordet bok betydde opprinnelig ei treplate av bøk som ble brukt til å risse runer inn i. I Bibelen betydde ei bok en papyrusrull med skrift på. De gamle islandske mestre skrev sine bøker på skinn. De eldste bøkene vi kjenner fantes antakelig først bare som muntlige overleveringer fra forteller til forteller, før de endelig ble nedtegnet i skriftlig form. Forestillinga om boka er altså i høy grad et produkt av teknologi.

Leksikografien er en bestemt disiplin innen vitenskapene om ord. Den har til de grader vært bundet til den trykte boka at den har tatt preg av det. Den trykte bokas omfang og begrensninger har virket tilbake på leksikografien og leksikografene. Nå som den elektroniske boka er på full fart inn i hverdagen, vil vi om litt bli konfrontert med dette.

1 Den trykte bokas begrensninger

Den trykte boka har satt klare grenser for

- omfanget av stoffet
- systematikken
- rekkefølgen
- presentasjonsformen

La oss se på disse begrensningene:

1.1 Omfang

Den norske *Bokmålsordboka* har ca. 5000 tegn pr. side på i alt 697 sider, dvs ca. 3,5 mill. tegn. I forhold til en standard roman på kanskje vel en halv million tegn, er jo ordboka reine mursteinen. Men i forhold til kapasiteten på en CD-ROM, blir mursteinen redusert til en grusbit, eller nærmere bestemt en halv prosent av de 700 millioner tegn ei slik plate kan romme. Til sammenlikning er det som om det bare skulle vært trykk på de første 3,5 sidene i *Bokmålsordboka*, mens de resterende 693 sidene var blanke.

Ønsket om å klemme maksimalt med ordboksinformasjon inn i ordbøkene har skapt et eget ordbokspråk med forkortelser, komprimerte setninger, koder og eventuelt også symboler. I *Bokmålsordboka* kan f.eks. en artikkel se slik ut:

I bo subst (Ity *behof*, sm o s *behov*) i uttr *ha bo* trenge, ha bruk for

I denne korte artikkelen er det seks forkortelser og en kode. Hadde ikke ordboka vært skrevet på denne måten, ville den kanskje krevd tre ganger så mange sider. På CD-ROM forsvinner dette argumentet. Vi ville fortsatt ha hatt 98,5% av plata til disposisjon.

1.2 Presentasjonsform

Den trykte bokas fysiske begrensninger påvirker også presentasjonsformen. Hvor mye grammatikk bør ei ordbok inneholde? Hvor mange eksempler? Hvordan skal sammensatte ord framstilles?

Vi har vendt oss til å bli avspist med kode- og telegramspråk. Bare den aller nødtørftigste grammatikk får plass og da gjerne i form av koder som *m1*, *v2* og *a5*. Eksempelene er sjelden belagt med formuleringer fra det virkelige liv, hvis vi da ser bort fra svære verker som *The Oxford English Dictionary* i nærmere 20 bind. Vil vi ha etymologi ut over det rudimentære, må vi gå til egne etymologiske ordbøker. I norske ordbøker settes gjerne sammensatte ord inn under en hovedartikkel, slik som

ord n1 (norr. . .) . . . **-bok** alfabetisk ordnet samling av ord fra et språk

Men med den plassen vi får til rådighet på en CD-ROM kan vi gjøre dette radikalt annerledes.

I en del ordbøker og de fleste leksika brukes illustrasjoner for å forklare teksten bedre. Den første som utstyrte et oppslagsverk med illustrasjoner skal ha vært den hellige Isidore fra Sevilla. Denis Diderot la meget stor vekt på illustrasjonens betydning i den store franske encyclopedien. Tekst og illustrasjon skulle bidra til å sprengte laugenes kunnskapsmonopol og bringe kunnskapen ut til folket. Slik fikk illustrasjonen en viktig rolle i encyclopediske verk. Men fortsatt har den trykte boka tvunget oss til å holde oss til skriftlig framstilling av lyd og bevegelse. Med elektroniske leksika stiller saken seg ganske annerledes.

Jeg skal komme tilbake til noen forslag.

1.3 Rekkefølge

Den vanligste rekkefølgen i ei ordbok er alfabetisk orden. Men vi kjenner mange avvik fra dette. I kryssordbøkene er betydning og antall bokstaver som regel overordnet alfabetet. I rimordbøker er slutten av ordet viktigere enn begynnelsen. I tesauriske oppstillinger kan fagområde være første nøkkel.

Men felles for alle mulige rekkefølger i den trykte boka: *når rekkefølgen er valgt, ligger den fast*. Den trykte boka er et *lineært* medium. Det er ikke den elektroniske boka. Her er alle ønskelige rekkefølger mulige. Utfordringa for leksikografen blir da delvis å ha tenkt gjennom hvilke alternative sorteringsmuligheter som skal kunne tilbys og delvis å sørge for at innholdet i den elektroniske boka er slik at den alternative sorteringa gir mening. Hvis alle oppslagsord er forsynt med ordklasse, kan f.eks. brukeren sortere dem først etter ordklasse og dernest omvendt alfabetisk, så har hun en rimordbok over adjektiver!

Mange ordbøker og de fleste leksika bruker henvisning til andre del av samme verk. For brukeren kan dette bli tungvint, så vi prøver derfor å redusere slik kryssreferanse til det absolutt nødvendige. I elektroniske verk stiller dette seg også helt annerledes. Intelligent

krysshenvisning, såkalt *hypertekst* blir både enkelt og morsomt i bruk. Men det stiller nye krav til leksikografen.

1.4 Systematikk

I encyclopediske verk er systematikken helt sentral. Den kommer ofte ikke fram i presentasjonen, men ligger bak kunnskapsmengden som en skjult struktur. Den encyclopediske tradisjonen går tilbake til middelalderens *Summae*, og graver man lenge nok, vil man finne spor av denne arven i encyclopedienes skjulte struktur. Så lenge encyclopedien var en sidedisiplin til teologien, finner man teologenes verdensoppfatning meget klart uttrykt i systematikken. Vincent de Beauvais delte sin *Speculum majus* opp i tre deler: *Naturale* — Gud, skapelsen og mennesket; *Doctrinale* — språk, etikk, fagene, medisin; *Historiale* — verdenshistorie. Francis Bacon tok skrittet ut av denne teologiske tradisjonen da han delte opp sin *Instauratio magna* i følgende tre deler: Den *ytre naturen* (omfattet bl.a. astronomi, geografi, meteorologi), *mennesket* (anatomi, fysiologi osv.) og *menneskets påvirkning av naturen* (bl.a. medisin, kjemi, arkitektur og transport og en mengde andre emner). Den store franske encyclopedisten Denis Diderot overtok og videreførte Bacons systematikk og gjennom ham har denne arven blitt ført videre til oss. Denne systematikken har sine utvilsomme styrker, men den har også innebygd i seg en del svakheter. Det jeg særlig vil legge vekt på her er faren for å utelukke tverrfaglighet. Verden splintres opp i enkeltvitenskaper for at den skal bli lettere å katalogisere og beskrive, men den virkelige verden eksisterer bare i sammenheng og samhandling. Hvis vi griper den nye teknologien på en riktig måte, skal vi kunne gjøre interessante forbedringer av oppslagsverkene fra en slik synsvinkel.

2 Noen forslag til hva leksikografer kan bruke CD-ROM til

Jeg begynte å arbeide med leksikon på CD-ROM i 1986, samme år som teknologien ble lansert internasjonalt og fikk æren av å lansere Europas første leksikon på CD-ROM i april 1987. I disse dager lanserer vi det tredje CD-ROM-leksikonet som jeg er redaktør for. Nå med lyd, interaktive kart, massevis av bilder, video, animasjoner og lyd. Disse erfaringene har lært meg en del om hvilke muligheter dette mediet inneholder for oss leksikografer. Henning Bergenholtz (1995) sa i sitt foredrag at elektroniske ordbøker til nå bare har vært elektroniske versjoner av de trykte bøkene. Dette kan jeg skrive under på. Ingen har ennå tatt inn over seg til fulle de endrede vilkårene for leksikografien som de nye mediene gir. Derfor vil jeg benytte denne anledningen til å utfordre kremen av nordiske leksikografer til å være med å forme en leksikografisk teori og framstillingsform som tar elektronikken på alvor. Jeg vil skissere noen av dem og peke på de teoretiske utfordringene som følger i kjølvannet av disse mulighetene. For å sette fantasien i sving har jeg tillatt meg å beskrive ei stor norsk encyclopedisk ordbok som ikke er utgitt eller planlagt. Vi forestiller oss at dette er både ei ordbok og et leksikon, slik at vi kan få nytte av alle de mulighetene som mediet gir.

2.1 Omfang

Vi begynner med omfanget:

Siden vi har så mye plass, bør vi legge opp til å bruke den. Vi begynner derfor med å sløyfe forkortelser og koder. Vi skriver inn den fulle bøyningstabellen for alle oppslagsord. Dette gjør vi ikke bare for å bedre innholdet i den enkelte artikkelen. Men dette gjør at alle former av ordet vil være søkbar i databasen. Et søk i fri tekst etter ordet *ordbøker* vil gi treff selv om oppslagsordet er *ordbok*. Av samme grunn må vi slutte å føre opp sammensatte ord på den tradisjonelle kortformen: *ord, ... -bok*. Allerede her vil vi merke at formen på artikkelen vil endre seg.

Så bestemmer vi oss for å legge inn et par norske grammatikker. Vi kople dem intelligent til hver enkelt ordboksartikkel, slik at brukeren kan få øyeblikkelig grammatikalsk hjelp. Her må leksikografen leve seg inn i brukerens situasjon og tenke ut hva slags støtte hun vil ønske.

Berit Heer Henriksen (1995) ga en lærerik framstilling av problemene knyttet til lemmatisering av avledete begrep, sublemma osv. Spørsmålet er hvor mye av dette som knytter seg til papiret som begrensning. Plassen hindrer oss ikke lenger i å lage det antallet sublemmaer vi måtte ha bruk for. Og det stående uttrykket kan utmerket godt brukes som sublemma for så mange ord i uttrykket vi måtte ha behov for. Søker du på *barn*, finner du *brent barn skyr ilden*, blant mange andre eksempler. Men du kan også finne det dersom du har søkt på *ild*. Leksikografens problem er hvordan disse mulighetene skal tilrettelegges for å gi brukeren mest mulig nytte og glede. For eksempel bør det finnes en mulighet for å hente fram alle idiomer, talemåter og stående uttrykk, samlet og hver for seg.

Siden plass ikke er noe problem for oss, så legger vi inn både nynorsk og bokmål på samme plata. I neste versjon bestemmer vi oss for at vi også skal legge inn ei samisk-norsk ordliste, slik at man kan slå opp betydninga av enkeltord begge veier. For å være forberedt på dette, har vi alt i første versjonen tenkt ut hvordan dette skal gjøres.

Ennå har vi ikke brukt 5% av plata, så vi kan fortsatt boltre oss en masse. Vi legger inn ei svær eksempelsamling fra skjønnlitteratur, aviser og dagligtale, komplett med årstall og kilde. I tillegg til det arbeidet språkfolk allerede har gjort på dette området, vil vi ha muligheten til å bruke de store avisdatabasene for å finne gode eksempler på bruk av bestemte ord. På denne måten blir den generelle ordboka også ei sitatordbok og en språkhistorisk dokumentasjon. Kanskje vi også ønsker at brukerne skal få tilgang til svære korpussamlinger. Da legger vi inn en valgmulighet for å gå fra den elektroniske ordboka og over til en stor database via Internett.

Noen leksikografer, eller kanskje mest encyclopedister, hvis vi får bruke et slikt ord, har eksperimentert med å føre inn lydlike oppslagsord med henvisning til riktig skrivemåte. Det ble f.eks. gjort i *Damms store leksikon* i Norge, men har ikke slått gjennom. I vår digitale ordbok kan vi åpne muligheten for såkalt *soundex*-søk, dvs. at vi søker på lydlikhet. Hvis vi taster inn *fono*, vil vi også finne *phono*. I motsetning til i den trykte boka, er dette ikke noe plassspørsmål lenger, det er et programmeringsspørsmål. Apropos programmering: når vi først er i gang, bør vi sørge for at boka er programmert slik at man kan få tilgang til den fra et tastatur for blindeskrift. Da vil vi plutselig ha ei svær encyclopedisk ordbok for blinde samtidig som for seende.

2.2 Systematikk

Så har vi systematikken:

Ordklassene har vi sjølsagt med. Dette vil glede for eksempel kryssordløserne, fordi de nå lett kan finne "preposisjon på syv bokstaver der fjerde bokstav er *n*". Men vi bør også være nøye med å få med hvor faguttrykkene hører hjemme. På denne måten kan vi for eksempel finne fram alle maritime faguttrykk og glede oss over hvor mange av dem som kommer fra nederlandsk og plattysk. I leksikondelen av denne ordboka vil vi gå enda lenger i oppdeling av stoffet i fagområder. Der vil jeg også foreslå at vi tar med en thesaurus med synonymmer og relevante emneord.

Referanse til andre artikler legges inn som hypertext, slik at brukeren bare kan peke og klikke seg fram. Hvis vi indekserer hele teksten på riktig måte, vil brukeren ganske enkelt kunne peke på et hvilket som helst ord og finne hovedartikkelen for dette ordet.

Jeg kan ikke forlate dette emnet uten å ta med min favorittsystematikk. Den er hentet fra et kinesisk leksikon fra keisertida og inneholder følgende inndeling av dyreartene:

- dyr som tilhører keiseren
- balsamerte dyr
- tamme dyr
- pattegriser
- havfruer
- fabeldyr
- frittgående hunder
- dyr som beveger seg som sinnsyke
- utallige dyr
- dyr som er tegnet med tynn kamelhårspensel
- dyr som nettopp har unngått fare
- dyr som ligner fluer på avstand

2.3 Søking i fri tekst

Vi legger sjølsagt opp til at brukeren skal kunne søke både i fri tekst og etter fagområde, ordkombinasjon og ordklasse. Kanskje bør det også være mulig å spesifisere antall tegn og kanskje også antall stavelser. Søking med synonymmer er også en mulighet vi bør vurdere. Siden vi har lagt inn ei synonymordliste, er det enkelt å programmere denne funksjonen.

2.4 Lyd

Først nå beveger vi oss fra tekst til lyd:

Nå er det ikke så få leksikografer som har fylt ordbøker med lyd, problemet er at det har vært skriftelig lyd, nemlig lydskrift. Det er jo en bleik erstatning for virkelig lyd. Vi har også det problemet at jo riktigere lydskriften er, fra en faglig synsvinkel, jo ferre er det som har kompetansen til å forstå den; og jo lettere den er å forstå, jo mer omtrentlig er den. Så nå som vi har muligheten til å gi brukerne våre den virkelige lyden i stedet for den skriftlige forflatninga av den, bør vi gi dem det. Tenk bare på hvor mye enklere det blir å forklare tonelagsforskjeller eller palatalisering, eller klikkelydene i xhosaspråket. Dialektpøver blir

også mer interessante på denne måten. Vi bestemmer oss altså for å la alle oppslagsord ha uttale og det samme med en del av eksempelmaterialiet.

Men lyd har også sin egen forklaringsverdi. I *Cappelens store musikkleksikon* ser definisjonen på en tritonus slik ut:

det intervall som dannes av tre heltonetrinn og som i den diatoniske skala av stamtoner representeres av intervallet f-(g-a-)h, dvs. en forstørret kvart.

Dette er forståelig for alle som ikke trenger å få begrepet forklart og uforståelig for alle som trenger det. Litt lyd her ville vært av det gode. Men dette er porten inn til en hel verden av lyd for ordbokfolket. I CAPLEX, Cappelens ettbinds leksikon, er fagotten forklart slik:

fagott, treblåseinstrument med dyp tone og nasal klang.

Fra nå av venter brukerne at vi også lar dem få høre hvordan instrumentet lyder. Hvis ikke leksikografen nå har skjønt at ordboka er i ferd med å bli noe helt annet enn vi er vant til, vil det gå opp for ham når vi bestemmer oss for også å inkludere bevegelse som enda en ny dimensjon i ordboka.

2.5 Illustrasjonen må med

I følge et gammelt kinesisk ordtak som oftest feilsiteres, er det bedre å se en ting en gang, enn å høre om den hundre ganger. (Kineserne ville aldri ha satt bildet opp mot ordet, både på grunn av sin respekt for ordet, og fordi deres ord er piktogrammer.) En teknisk innretning kan av og til beskrives tilfredsstillende ved hjelp av bare ord. Men de fleste har innsett at det blir enklere, hvis man også har en illustrasjon å støtte seg på.

Riksmålsordboka definerer en sjakkell slik:

sjakkell, bøile av metall som i begge ender har hull (øie) for en bolt som den lukkes med, brukt til å forbinde med hverandre (særl.) gjenstander som er forsynt med (ell. ender i) ring ell. bøile.

Vi kan formelig føle hvordan leksikografen sliter med å presse virkeligheten inn i språket. En liten skisse i tillegg ville gjort det ganske mye enklere. Dette var jo noe vår læremester Diderot innså til fulle, og i vår visuelle tid er det ikke blitt mindre viktig. Nå kan vi heller ikke skyldes på plassproblemer. Strekillustrasjoner av typen sjakkell kan vi antakelig legge inn ti tusen av uten å oppleve plassproblemer. Problemet blir i stedet leksikografens. Hun blir plutselig bilderedaktør også. Kanskje kan det være fornuftig å lage samleillustrasjoner, slik man gjør i illustrerte ordbøker, men med den forskjellen at vi kan indeksere hver enkelt bildetekst slik at den blir søkbar og lagt inn i kryssreferansetabellen.

2.6 Ordbok med bevegelse

Bevegelse i ordbok eller leksikon, helligbrøde, jåleri, vulgært. Hva skal vi med det? Det skal innrømmes at mye av det vi har sett av video på multimedia oppslagsverk har vært heller tvilsomt fra en faglig synsvinkel. Men det er nå en gang slik at virkeligheten ikke foregår bare i to eller tre dimensjoner. Den fjerde, tidsdimensjonen, er i høyeste grad en

realitet, noe i det minste våre grå hår burde minne oss om. Den reine ordlista har sjølsagt ikke særlig behov for å kunne vise bevegelser. Men med en gang vi beveger oss over til definisjonsordboka, for ikke å snakke om leksikonet, så ser vi at vi svært ofte bruker tidsdimensjonen i skriftlig framstilling. Vi simulerer tid gjennom skrift, vi lager skriftlig bevegelse. Dette kan vi ha behov for når vi skal forklare så forskjellige begreper som platetektonikk, sluse, firetaksmotor, matematiske funksjoner, histogrammer og frontenes utvikling under anden verdenskrig. Man skulle kunne tenke seg at dette ville være en fin måte å følge opp det arbeidet Leena Savolainen (1995) fortalte om i forbindelse med ordbok for tegnspråk. Hun har imidlertid fortalt meg at dagens CD-ROM teknologi gir for lite plass til video. Men om fem år er det problemet løst, så vi kan like gjerne begynne å tenke i de banene.

3 Utfordringene blir altså

Til nå har jeg snakket om muligheter med denne nye teknologien. Hva har det med leksikografisk teori å gjøre? En av kollegene var så ufin at han spurte om det finnes en leksikografisk teori. Men la oss gå ut fra at det gjør det. De utfordringene jeg ser er at vi må begynne å utvikle begreper, framstillingsmåte, kriterier og metodikk som tar helt og fullt inn over seg at vi skal publisere for et elektronisk medium.

For at den elektroniske ordboka skal bli god, er god og hensiktsmessig formatering eller koding alt.

Spørsmål 1: Hva er god og hensiktsmessig formatering i denne sammenhengen?

Spørsmål 2: Gitt at vi har all denne plassen til rådighet, hvordan bruker vi den? Hvor går grensa mellom god og rikholdig informasjon og overflødig eller hemmende informasjon?

Spørsmål 3: Skal vi tilby brukerne alternative fagkodesystemer? Hvilke er disse og hvordan skal de presenteres?

Spørsmål 4: Hvilke andre måter kan vi utnytte ikke-lineariteten på? Synonymordbok, enderimsordbok, sitatordbok, fagordbok osv. i samme verk. Er det mulig å gjøre dette på en god måte, og i tilfelle hvordan?

Spørsmål 5: Det finnes muligheter for å tilby brukeren forslag til søkestrategier, som hun kan bygge videre på. Hvilke bør brukes og hvordan skal de tilrettelegges. Hva med hjelpeverktøyenes rolle? Skal brukeren få muligheten til å bruke konkordansverktøyer i ordbok/leksikonet vårt?

Spørsmål 6: Hvordan leksikaliserer vi lyd, bilde, animasjon og video? I billedarkiver lagres bilder med en beskrivelse av bildet, f.eks. om det er fra Island og viser en leksikograf som drikker "Svarti dauði" foran en vulkansk fjellvegg. Skal de nye mediene beskrives på denne måten, slik at innholdselementene i dem blir søkbare og hvordan skal det gjøres? Når blir slike medier verdifulle tilskudd til verket og når blir de hype? Skal det legges inn tabellarisk informasjon og skal den kunne manipuleres?

(Skal for eksempel folketellingsdata kunne tas inn i et regneark og legges ut som grafer?)

Så velbekomme, kjære kolleger leksikografer. En ny generasjon brukere av ordbøker og leksika er kjent med hvilke muligheter de nye mediene gir. De venter av oss at vi skal bruke disse mulighetene.

Litteratur

- Bergenholtz, Henning 1995: Leksikografi. Hvad er det? [Denne rapporten, 37–49.]
- Bokmålsordboka = M. I. Landrø/B. Wangensteen 1993: *Bokmålsordboka. Definisjons- og rettskrivningsordbok*. Oslo: Universitetsforlaget.
- Cappelens musikkleksikon* 1978–80. Oslo: Cappelen.
- CAPLEX. Cappelens ettbinds leksikon* 1991. Oslo: Cappelen.
- Damms store leksikon* 1985–90. Oslo: N. W. Damm & søn.
- Henriksen, Berit Heer 1995: Korpusbaserte relationsopplysninger og lemmatisering af flerordsforbindelser. [Denne rapporten, 195–203.]
- Riksmålsordboka = *Norsk riksmålsordbok* 1937–57. Oslo: H. Aschehoug & Co.
- Savolainen, Leena 1995: Finnish Sign Language-Finnish dictionary: defining the head signs. [Denne rapporten, 359–364.]
- The Oxford English Dictionary* 1989. I–XX. Second Edition. Oxford: Clarendon Press.