

Nettbrukaren i fokus: digitalisering og revisjon av Norsk Ordbok *a–h*

Johan Myking, Oddrun Grønvik & Stig Jarle Helset

The dictionary Norsk Ordbok (Norwegian Dictionary) was initiated in 1929 and accomplished in 2015 with the publication of its twelfth and final volume. The Dictionary covers the Nynorsk written standard as well as the vocabulary of Norwegian dialects in all their variety. Two thirds of the 330 000 lemmas (letters *i–å*) are open to digital search. Work has now been initiated to revise the contents of the first third (letters *a–h*) as well as making this part available for digital searches. The revision will have to adapt to the needs of digital-minded user groups, as no updated paper version is planned. Such adaptive measures include the extensive use and development of electronic corpora.

1. Innleiing

Norsk Ordbok er den einaste ordboka som startar i tidleg nynorsk periode (rundt 1600) og dekkjer heile tidsspennet fram til i dag. Den er den einaste norske ordboka som kan gjera vitskapleg bruk av den komparative språkvitskapen til å identifisera norsk ordtilfang diakront og synkront. Meir enn halvparten av ordtilfanget i Norsk Ordbok er uttrykt der i normert form for første gong.

Arbeidet med Norsk Ordbok vart sett i gang i 1929 ved Universitetet i Oslo og vart avslutta i 2015. Det Norske Samlaget har gjeve ut verket i 12 band, fullført gjennom prosjektet Norsk Ordbok 2014 (NO2014) i åra 2002–2015. Verket inneheld om lag 330 000 oppslagsord. Dei siste to tredelane, bokstavstrekket *i–å*, er digitalt tilgjengelege.

Ordboksverket inkluderer eldre kjelder og siktar mot å dokumentera leksikalsk talemålsvariasjon i eit større omfang enn vanleg i ordbøker der litterære kjelder dominerer. Norsk Ordbok fører vidare Ivar Aasens prinsipp for skriftstandardisering på talemålsgrunn:

Den gjennomgåande dekninga av norsk talemål sikrar at alle brukarar av norsk skal kunna finna att noko av sitt språk i verket, systematisert, vitskapleg fortolka og løfta inn i ein samheng som viser både dialektale særdrag og språkleg samheng mellom ulike område av landet. (Wetås & Aa 2017:291)

Ordboksarbeidet ved Universitetet i Oslo vart fasa ut og avslutta då papirutgåva av Norsk Ordbok var ferdigstilt. Materialet og det faglege ansvaret vart flytta til Universitetet i Bergen i 2016 og etablert som Språksamlingane ved Universitetsbiblioteket der, saman med andre delar av dei samlingane av språkleg materiale som Universitetet i Oslo sa frå seg ansvaret for. Tre typar infrastruktur blir drøfta nedanfor: Ordbokshotellet, Metaordboka og Nynorskorpuset, sjå avsnitta 4 og 5. Dessutan er Setelarkivet viktig, der ligg 3,2 millionar ordsetlar samla inn frå 1930-talet og til i dag lagra i digital form. Denne infrastrukturen utgjer hjelperessursane som var utvikla for Norsk Ordbok, men blir altså vidareført i Språksamlingane.

2. Ideen om Norsk Ordbok som nettordbok – NO-AH

Universitetet i Bergen (UiB) har i samarbeid med Høgskulen i Volda (HVO) teke ansvaret for å gjera ferdig den digitale utgåva av Norsk Ordbok og revidera alfabetstrekket *a–h*. Prosjektet har fått akronymet NO-AH, og har etter søknad til Kulturdepartementet fått midlar over statsbudsjettet, i første omgang for 2019 og 2020. Mange av dei faglege føresetnadene peikar utover alfabetstrekket *a–h*, men det er altså dette strekket prosjektet må konsentrera seg om.

NO-AH blir eit nytt leksikografisk prosjekt som må retta seg mot nettbrukaren og derfor byggjast på til dels nye faglege prinsipp, av di det ikkje kjem noka ny papirutgåve. Ideen om overgang til digital publisering kom som del av den omfattande digitaliseringa av produksjonsprosessen etter 2002:

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. (Ore 2012:489)

I søknaden til Kulturdepartementet for 2019 er målet for prosjektet formulert slik: «fullføra Norsk Ordbok på nett ved å revidera innhald og form på alfabetstrekket *a–h* (NO-AH), og slik skapa ei komplett elektronisk Norsk Ordbok». Det planlagde resultatet er då eit elektronisk system av ordboksdatabasar i Språksamlingane som skal gje ein eksplisitt, lettbrukt og vitskapleg fundert presentasjon av norsk ordtilfang.

2.1. Ein brukaranalyse i fem drag

Utviklinga av ei slik moderne digital ordbok må byggjast på ein brukaranalyse, fordi ordbokbrukarane i dag kan ha utvikla andre føresetnader for bruken av ordboka enn dei som galdt då første band kom ut i 1966. Vi ser såleis fem grunnleggjande drag ved det norske publikumet som må få følgjer for arbeidet:

- 1) Det norske publikumet har svært stor interesse for språk, både for tale, skrift og etymologi

Informasjonsbehovet må tilpassast det allmenne språkinteresserte publikumet, som er stort, og derfor må materialet òg vera så heildekkjande som mogeleg. I og med at dei digitale kunnskapane varierer, må ein då ha ei vurdering av brukarkompetanse: Kor stor språkkunnskap, digital kunnskap og kunnskap om ordboksstruktur kan ein venta av publikum, og kva konsekvensar skal det få? Dette er av dei spørsmåla som vil bli greidde ut tidleg i prosjektperioden, i samråd med representantar for viktige brukargrupper.

- 2) Det norske språksamfunnet er karakterisert av stor variasjon, både i tale og skrift

Ordboka må derfor førast vidare ut frå det etablerte prinsippet om å inkludera både talemål/målføre («det norske folkemålet») og skriftspråket, med eit dokumenterande meir enn normerande formål:

Eit ordbokverk med eit eksplisitt dokumenterande føremål og med ei diakron kjeldedekning som strekkjer seg over ein periode på 500 år, har eit anna føremål og eit anna potensielt publikum enn ei preskriptiv produksjonsordbok (Wetås & Aa 2017:291).

Den normerande effekten kan ein dels oppnå med koplingar til dei to offisielle standardordbøkene Nynorskordboka og Bokmålsordboka, dels ved at oppslagsorda er kopla til gjeldande rettskrivingsformer. Eit grensesnitt mot forskning i norsk talemål kan sikra at variasjonen i talemålet blir synleggjord (jf. målføresynopsisen, nedanfor).

3) Det norske samfunnslivet er prega av aukande og intensivert digitalisering og «snakkande» teknologi

I søknaden om prosjektet *Norsk ordbok a–h* heiter det at resultatet av arbeidet ikkje berre skal vera ei ordbok, men eit system av databasar og språkressursar. Desse ressursane må ha gjenbruks- og overføringsverdi, dei må vera utvikla med tanke på integrasjon i språkteknologi som del av dagens satsingar på digital infrastruktur.

4) Papir er «ute»

Dette er eit trivielt faktum: Det digitale mediet overtek, ingen nye papirutgåver er p.t. planlagde av nokon av dei større norske ordboksverka, som NAOB og standardordbøkene. Samstundes veit ein at brukargruppa for Norsk Ordbok ikkje er homogent datakompetent – ein skal leggja til rette både for dei som er mest vane med papirmediet, samstundes som praksisendringar gjev høve til å tilfredsstilla andre og meir avanserte brukarbehov enn i dag.

5) Eldre band treng revisjon!

Den siste føresetnaden gjeld ikkje mediet, men innhaldet. Band I av ordboksverket kom ut i 1966, deretter kom det tre band i 1978, 1994 og 2002 før ein gjekk over til digitale produksjonsformer i gjennom prosjektet NO2014 frå 2002. Dei fire første banda og første halvdel av band V er ikkje digitalt tilgjengelege. Verda har endra seg i dei meir enn 50 åra sidan band I kom ut, dels ved at nye omgrep er komne til, dels ved at eldre fenomen blir sedde gjennom nye briller. I band I finn ein følgjande artikkelar om dei opphavlege innbyggjarane i Australia:

australiemann m. [-ra'] (He.Fo, Vo.Ty), australiar m [-ra'-; til *Australia*, avl. t lat *australis* 'sørleg' av *auster* 'sønnavind'] (kvit) person som eig heime i el er frå Australia. [...] **australneger** m [-ra'] ein av det opphavlege folket i Australia.

Nemninga *aborigin* kom i bruk sist på 60-talet, og vil bli redigert i den reviderte Norsk Ordbok, men dei eksisterande artiklane kan ikkje sløyfast. I den digitaliserte versjonen vil denne typen oppslagsord bli merkte som forelda, medan definisjonane vil bli justerte.

2.2. Praksisendring og endra vektlegging

Som nemnt kan overgangen til det digitale mediet krevja endra leksikografisk praksis på nokre område, som skal presenterast nedanfor. Samstundes skal ein ta vare på heilskapen i verket, og må dermed respektera dei etablerte konvensjonane så langt råd er. Verket skal helst framstå nokolunde samla og homogent i form og struktur.

Når det gjeld innhaldssida, er det rimeleg å ta omsyn til eit primærbehov hos moderne nettbrukarar, som er å finna svaret på spørsmålet med ein gong. Dette gjeld særleg definisjonar og bruksdøme. *Norsk Ordbok i-å* (frå band V) vart redigert både for papir og nett, og redigering for papir inneber plassparing. Norsk Ordbok har derfor mange snautt formulerte definisjonar. I NO-AH vil det vera meir rom for fullskrivne definisjonar. Der *Norsk Ordbok i-å* måtte nøya seg med synonymdefinisjonar, vil NO-AH kunna kosta på seg både ein fleirordig definisjon og ei synonymtilvising eller to.

I kva grad desse ulike omsyna kan harmoniserast, vil visa seg eit stykke inn i framtida. Fokuset no ligg på alfabetstrekket *a-h*, men det vil opplagt bli behov for innhaldsrevisjonar av strekket *i-å* òg, sjølv om dette ved dags dato ikkje ligg inne i søknadene om finansiering.

3. Korleis presentera talespråkleg variasjon?

Som alt nemnt, har den grunnleggjande ideen bak Norsk Ordbok vore å gje ein dokumentasjon både av det nynorske skriftspråket og av det talemålstilfanget som ligg til grunn. Dermed har Norsk Ordbok ei større utfordring enn mange andre ordboksverk: Korleis bør informasjon om målførevariasjon presenterast i ei vitskapleg nettordbok, slik at utbyttet blir optimalt for heile spennet av ordboksbrukarar?

Den behovsanalysen – altså kva ein går ut frå at brukarane treng å få vita – som ligg til grunn for målføreinformasjonen i Norsk Ordbok, er ikkje direkte uttrykt, men framgår av opplysning om at føreseielege (sjølvgeivne) målføreformer er sløyfa. Norsk Ordbok på papir vender seg primært til lesarar som kan systematikken i det norske målførekartet. Omtalen i brukarrettleiinga (Vikør 2018:78ff.) syner at 1) oppstillinga av former (med kjelde) føreset kunnskap om konvensjonane for bøyingsmorfologi i norsk, og dermed litt meir enn grunnleggjande grammatisk terminologi; 2) dei viktigaste isoglossane for norsk talemål er føreset kjende, og likeins det områdekartet som desse isoglossane gjev opphav til. Såleis er omtalen av jamvekt, «tjukk l», infinitivsending og apokope

ofte forenkla til det som blir kalla «samleformer». Det gjer at forma «skjere-» i dømet nedanfor også blir brukt til å representere jamningsformer i tilfanget, som /sjyrru-/ og /sjurru-/:

I **skjere-** i sms [målf òg (stundom berre i einskilde sms) *skjer-* (Furnes, Folldal, Gbr sumst, Hall, Uvdal, Ri, Modum, Andebu, Tinn, Rog sumst, Selje, Bud, Ndm og Trl og No sumst), *skjera-* (SAustl sumst, Jr, Vik, Saltal, Beiarn, Gildeskål), *skjerar-* (YSandsvær, Andebu, Rødøy, Meløy, Kjerringøy, Vstrål, Vadsø); jfr òg *skjerarfolk* og *skjerarlår*] til *II skjera*. 1 om handlinga å skjera, såleis t d: skjere-fart, -måte, -resultat, -rørsle. 2 om (del av) reiskap til å skjera med, såleis t d: skjere-kant, -maskin, -reiskap, -verktøy. 3 om underlag til å skjera noko på, såleis t d: skjere-brett, -fjøl. (Transkribert versjon, redigeringshandboka (Grønvik m.fl. (red.). 2016: 162)

I dømet ovanfor kan ein røynd brukar av Norsk Ordbok sjå at poenget med målføreoppstillinga er å visa fugevariasjonen for førsteleddet, fordelt geografisk. For nye brukarar av Norsk Ordbok vil dette ikkje vera opplagt. I nettutgåva kan ein også få vist dokumentert utbreiing i kart, noko som raskt vart populært då tenesta vart lansert i 2014.

Ei revidert Norsk Ordbok vil som før sagt vera bunden av eksisterande konvensjonar. Men dagens ordboksbrukarar skal også kunna ha nytte og glede av målføreopplysningane utan så omfattande førehandskunnskapar som tidlegare føresett. Det står att å finna ut kva slags form på talemålsinformasjonen eit breiare publikum ser seg tent med. Det må uansett bli eit system som kan gjera bruk av eksisterande oppsett i redigerte ordartiklar.

4. Det digitale etterslepet etter NO2014

4.1. Korleis handtera eldre kjelder?

Noverande band I er tynt når det gjeld dei eldste kjeldene, men denne mangelen kan no supplerast gjennom bruk av dei digitale verktøya som har vore utvikla for Norsk Ordbok, særleg den såkalla Metaordboka, men òg andre. Viktig digitalisering er gjennomført i tida etter at NO2014 vart avslutta – særleg er det viktig at ordbøkene til Aasen (1873) og Ross (1895) er digitaliserte og no ligg i det såkalla Ordbokshotellet (jf. Grønvik & Ore 2018), slik at dei er søkbare i seg sjølve og via Metaordboka. At NO-AH blir heildigital, vil innebera at redaktørane får tilgjenge til det digitale etterslepet etter NO2014, som t.d. fulldigitalisering av talemålskjeldene Norsk målføresynopsis og Norsk dialektatlas, og nettilgjenge til dei eldre ordsamlingane (1600–1850).

I Oslo hadde redaksjonen tilgjenge til fleire unike papirarkivaliar, men det må det no bli slutt på – risikoen og slitasjen er for stor. Heildigitalisering må innebera at Setelarkivet berre blir digitalt tilgjengeleg, slik at ein ikkje lenger sorterar lappar på skrivebordet. Dermed må ein gjera noko med dei 18 000 setlane som manglar registrert oppslagsord og dei ca. 120 000 setlane som manglar registrert kjelde (synleg på foto, men ikkje registrert i metadata).

4.2. Metaordboka

Metaordboka er eit register over alt tilfang som er knytt til ein gjeven «leksikalsk gjenstand» ('lexical item', jf. Atkins & Rundell 2008), dvs. ein teiknstreng som krev eigen ordboksartikkel. Metaordboka knyter (knippe av) daterte oppslagsformer til kvar artikkel, slik at den leksikalske gjenstanden får kopla saman sine normerte uttrykk over tid med tilfanget som utgjer grunnlaget for innhaldsanalysen:

Metaordboka er et register der hver innførsel eller «artikkel» har et hode som tillater indeksering av belegg med (1) språkform (bokmål og nynorsk), (2) flere oppslag for hver språkform, (3) normeringsstatus for hver oppslagsform, (4) datering av hver status, (5) ordklasse, (6) leddanalyse. (Grønvik & Ore 2018:89)

Det er meininga at ordboka skal bli ein felles indeks for Språksamlingane. I høve til Norsk Ordbok er Metaordboka reiskapen som skal gjera det mogleg å søkja i Norsk Ordbok i moderne rettskriving. Bakgrunnstilfanget for normering og datering ligg i ordlistene i Ordbokshotellet. Språksamlingane samarbeider med Språkrådet om den såkalla Ordbanken, som inneheld alle normerte norske ord med bøyingsformer. Ordbanken er ein fullformsgenerator med lenkjer til alle oppslagsformer i dei to standardordbøkene Bokmålsordboka og Nynorskordboka (Grønvik & Ore 2013). Om Metaordboka blir kopla til Ordbanken, vil ein også kunna opplysa om gjeldande bøying for oppslagsformene i Norsk Ordbok (utover 1938-rettskrivinga, som ordboka opplyser om). Metaordboka kan på denne måten utviklast til eit omfattande register over historisk normering.

5. Korpus – kjeldekritikk og kvalitet

5.1. Korpusgrunnlaget for Norsk Ordbok

For nynorsk er det bygd opp eit tekstkorpus med ein diakron dimensjon – 1866 til i dag, det såkalla Nynorskkorpuset. Eit tilsvarande for bokmål finst ikkje. Det er enno ikkje avklåra kort stort omfang korpusarbeidet kan få i NO-AH, men det er ein intensjon om sterk utvikling i regi av Språksamlingane. Norsk Ordbok og prosjektet NO-AH vil stå sentralt, men arbeidet må peika lenger – ut over prosjektperioden for NO-AH.

I tida før prosjektet NO2014 som starta i 2002, var parolen å utnytta alt tilfang til siste komma, men under NO2014 vart ordsamlingane brukte både som redigeringsgrunnlag og til å proporsjonera ordboka. No er overvekta av nytt avistilfang eit problem på grunn av mengda. Ein balansert korpuspolitikk peikar seg dermed ut som avgjerande metodisk grunnlag for det vidare redigeringsarbeidet.

5.2. Nynorskkorpuset

Nynorskkorpuset omfattar om lag 106 millionar ord frå tekstar som høyrer til ulike sjangrar, skjønnlitteratur, avistekstar, og sakprosa frå 1870-åra og fram til i dag. Etter ein pause på nokre år er arbeidet med utvikling av Nynorskkorpuset sett i gang att, med juridisk og fagleg grunnlagsarbeid. Korpuset har ein sjølvstendig bruksverdi i tillegg til den leksikografiske. Særleg er Språkbanken ved Nasjonalbiblioteket som leverandør av strukturert tekstmateriale til språkteknologiske formål ein stor interessent og mottakar av denne typen tekst.

Ved universitetet i Oslo vart det òg utvikla eit leksikografisk korpus over bokmål, og begge desse korpusa høyrer til Språksamlingane ved UiB i dag. I samarbeid mellom fleire universitet og fagmiljø er det òg utvikla eit stort Norsk aviskorpus. Potensialet er altså stort for å driva korpusbasert leksikografi på norsk.

Oppsummert vil dette seia at korpusdelen av arbeidet med NO-AH må ha tre mål: For det første må korpuset aukast og utvida så det blir meir representativt, eventuelt supplert med nye spesialkorpus. For det andre er leveranse til språkindustriell bruk eit politisk ønske som det er viktig å ta omsyn til. Og for det tredje, for å få dette til, må det leksikografiske fagmiljøet arbeida med auka vekt på kjeldekritikk og brukarkompetanse.

6. Norsk Ordbok a–h som løysinga på ein politisk situasjon

Norsk Ordbok har sidan den første starten i 1930 vore ei politisk sak på den måten at finansieringa har skjedd over statsbudsjettet. Ikkje minst var den radikale omlegginga til NO2014 i 2002 avgjerande for at verket låg føre i tolv band i 2015. Også spørsmålet om digitalisering og revisjon av strekket a–h er eit spørsmål som krev politiske løysingar. *Norsk Ordbok a–h* (NO-AH) er såleis resultatet av ein politisk situasjon og måten den vart løyst på.

Då UiB i samarbeid med HVO i 2016 gjekk inn og overtok Språksamlingane og ansvaret for digitalisering av *Norsk Ordbok a–h*, følgde ansvaret for dei to standardordbøkene Bokmålsordboka og Nynorskordboka med i porteføljen. Ein ser no at overtakinga føreset ei strategisk satsing på leksikografi som forskingsfelt. Ei slik satsing representerer eit potensial for utvida aktivitet og «gjødsling» av den nasjonale språkvitskaplege satsinga på norskspråkleg infrastruktur og språkteknologi.

Litteratur

Ordbøker

NAOB = *Det Norske Akademis ordbok*. Oslo: Det Norske Akademi for Språk og Litteratur og Kunnskapsforlaget. Digital utgåve: <www.naob.no>

Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet. Band I–XII. 1966–2015. Oslo: Det Norske Samlaget. Digital utgåve i-å: <no2014.uib.no>

Ross, Hans (1895): *Norsk Ordbog. Tillæg til 'Norsk Ordbog' af Ivar Aasen*. (1971). Fotografisk opptrykk. Oslo: Universitetsforlaget.

Aasen, Ivar (1873): *Norsk Ordbog med dansk Forklaring. Fjerde uforandrede udgave, (1918)*. Kristiania: Vestmannalaget og Cammermeyers forlag.

Annan litteratur

Atkins, Beryl T. (Sue) & Michael Rundell (2008): *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Grønvik, Oddrun m.fl. (red.) (2016): *Redigeringshandbok for Norsk Ordbok 2014*. <no2014.uib.no/eNo/tekst/lenkjer.html>

Grønvik, Oddrun & Christian-Emil Smith Ore (2014): Samvirket mellom ordbank og ordbok. I: Fjeld, Ruth Vatvedt & Marit Hovdenak (red.): *Nordiske studier i leksikografi 12. Rapport fra Konferanse om leksikografi i Norden Oslo 13.-16. august 2016*. Skrifter utgitt av Nordisk forening for leksikografi. Skrift nr. 13. Oslo. Novus forlag, 1–22.

- Grønvik, Oddrun & Christian-Emil Smith Ore (2018): Bokmål og nynorsk samindeksert – Metaordboka som verktøy for jamføring og utforsking av ordtilfang. I: Ásta Svavarsdóttir; Halldóra Jónsdóttir; Helga Hilmisdóttir & Þórdís Úlfarsdóttir (red.): *Nordiske Studier i Leksikografi 14, Rapport fra 14. Konference om Leksikografi i Norden Reykjavík 30. maj - 2. juni 2017*. Nordisk Forening for Leksikografi, 87–95.
- Ore, Christian-Emil Smith & Oddrun Grønvik (2018): Comparing Orthographies in Space and Time through Lexicographic Resources. I: Jaka Čibej, Vojko Gorjanc, Iztok Kosem Simon & Krek (eds.): *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*. Ljubljana: Ljubljana University Press, Faculty of Arts, 159–172. <e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/118/211/3000-1.pdf>
- Ore, Christian-Emil Smith (2012): Nettordbøker og Norsk ordbok – hvordan etablere en vitenskapelig nettordbok. I: Eaker, Birgot, Lennart Larsson & Anki Masttisson (red.): *Nordiska studier i Lexikografi 11. Rapport från Konferensen om lexikografi i Norden Lund 24.-27. mai 2011*. Skrifter utgivna av Nordiska föreningen för lexikografi. Skrift nr. 12. I samarbeite med Språkrådet i Norge og Svenska Akademiens ordboksredaktion, 488–499.
- Vikør, Lars S. (2018): *Inn i Norsk Ordbok. Brukarrettleiing og dokumentasjon*. <no2014.uib.no/eNo/tekst/lenkjer.html>
- Wetås, Åse & Leiv Inge Aa (2017): Museale betraktningar av eit moderne ordbokverk – tilsvar til John Ole Askedal si melding av Norsk Ordbok. I: *LexicoNordica*, 289–300.

Johan Myking
professor
Universitetet i Bergen
Boks 7805
NO-5020 Bergen
johan.myking@uib.no

Oddrun Grønvik
forskar
Universitetet i Bergen
Boks 7805
NO-5020 Bergen
oddrun.gronvik@uib.no

Stig Jarle Helset
førsteamanuensis
Høgskulen i Volda
Boks 500
NO-6101 Volda
stig.helset@hivolda.no