

Gamle ordbøker og digitale utgaver

Christian-Emil Smith Ore

Today it is relatively easy to re-publish old dictionaries on the web based on retro digitized text. To create trustworthy editions which can serve as the fundament for reliable Linked Data, this has to be done with care and it is necessary to follow the basic principles of digital scholarly text editions. A digital edition of the Norwegian Dictionary written by the Norwegian linguist Ivar Aasen published in 1873 is used as example.

1. Innledning

I denne artikkelen vil jeg diskutere en del forhold som jeg mener er av vesentlig betydning for digital utgivelse av gamle ordbøker. Utgangspunktet er det arbeidet som Oddrun Grønvik og jeg har gjort for å forberede en felles digital utgave av ordbøkene til de to norske språkforskerne Ivar Aasen og Hans Ross. Det er spesielt fire forhold jeg vil behandle:

En eldre ordbok er ikke en leksikalsk database i forkledning. Det gjelder i høy grad Aasens ordbøker. De er pionerordbøker med mye forklarende tekst.

Gamle ordbøker har en klar egenverdi som dokumentasjon av språket og kulturen på den tiden de ble skrevet, og er dermed gode verktøy til å forstå tekster fra denne tiden. Når eldre ordbøker utgis på nytt, må den originale teksten derfor presenteres fullstendig og uendret. Dersom en likevel ønsker å legge til tekst eller flytte på eksisterende tekst, må det være tydelig markert.

Papir og skjerm er to forskjellige medier. For en digital utgave er det hensiktsmessig å redigere ordboksteksten i mindre, selvstendige deler som kan leses uten kontekst. Eldre ordbøker har ofte et preg av å være skrevet som en sammenhengende tekst. For eksempel vil mindre ordartiklene for avledninger og sammensetninger ofte forutsette at en lett kan se hovedartikkelen for grunnordet. Det er verken god formidling eller innenfor god edisjonsfilologisk praksis å foreta en finoppdeling som bryter med ordboksverkets egenart.

De edisjonsfilologiske kravene og den eldre tekstens spesielle krav til visning er ikke til hinder for en dypere analyse og tilhørende oppmerking av

informasjonselementer som for eksempel normaliserte oppslagsord, ordklasse, stedsnavn, kryssreferanser og henvisning til andre verk. En slik oppmerking bør være obligatorisk siden den gir muligheter til å sammenstille opplysningene i mange ordbøker. Men analysen og oppmerkingen er en del av det moderne edisjonsfilologiske kommentarapparatet og ikke en del av det originale verket.

2. Aasens ordbok fra 1873 – posisjon, innhold og kategorier

I *Norsk Ordbog med Dansk Forklaring* som utkom i 1873 (Aasen 1873), legger den norske språkforskeren Ivar Aasen (1813–96) frem et utkast til et standardisert norsk skriftspråk basert på det samtidige talemålet. Ordboken bygger på Aasens omfattende innsamlingsarbeid i Norge. Aasens ordbok fra 1873 er et grunnleggende verk i studiet av norsk språk. Den bygger på hans *Ordbog over det norske Folkesprog* (1850) og på *Norsk Grammatik* (1864).

Filologen Hans Ross (1833–1914) fulgte opp Aasens innsamlingsarbeid og publiserte i 1895 ordboken *Norsk Ordbog, Tillæg til "Norsk ordbog" af Ivar Aasen*. Ross utgav seks supplement til sin egen ordbok i perioden 1895 til 1913. Ross' ordbok er redigert som et supplement til Aasens ordbok fra 1873 og må leses i sammenheng med denne.

Aasens ordbøker men også Ross' ordbok med tilleggene er svært viktige kilder for arbeidet med *Norsk Ordbok* (NO). De er ikke ekserpert i seddelarkivet til *Norsk Ordbok*, men de er tilgjengelige gjennom det såkalte grunnmanuskriptet (GMANUS) som finnes som en del av NO2014-prosjektets omfattende bakgrunnsmateriale. Grunnmanuskriptet er et maskinskrevet utkast til *Norsk Ordbok* fra 1935. En vesentlig del av dette manuskriptet utgjøres av ordbøkene til Aasen og Ross, der ordboksartiklene med de danske forklaringene er oversatt til norsk, tilpasset og flettet sammen til nye felles ordboksartikler. Grunnmanuskriptet inneholder altså en oversettelse og bearbeidelse av de originale tekstene til Aasen og Ross.

Det har lenge vært et ønske om å lage en elektronisk versjon av ordbøkene til Aasen og Ross. Men siden de har vært tilgjengelige gjennom Grunnmanuskriptet, de trykte utgavene og som digital brødtekst (se nedenfor), har en slik digital utgave ikke hatt høyeste prioritet i NO2014-prosjektet. Men det har like fullt vært arbeidet med å gjøre dem tilgjengelig i

databasene for NO2014-prosjektet. Tidlig i 2014 foreslo Oddrun Grønvik og jeg å lage et prosjekt for en digital utgave av Aasen og Ross' ordbøkene med vedlegg. Vi fikk klarsignal fra ledelsen i NO2014-prosjektet i 2014 og lagde en foreløpig versjon som var ferdig til mai 2015.

Listen nedenfor angir målene til hovedprosjektet vårt. For NO2014-prosjektet var punkt 1 nedenfor det viktigste. Det lenker ordbøkene inn i den felles kildedatabasen. Prosjektets mål er altså å

1. identifisere oppslagsord og lenke dem til moderne form gjennom Metaordboka, indekseringssystemet for språksamlingene
2. identifisere stedsnavn og tolke dem som en referanse inn i ett eller flere av herredene i 1947 og dermed georeferere referansene
3. finne sitater, målføreformer, og interne og eksterne referanser
4. gjøre alle ordbøkene tilgjengelige på nett og som Linked Open Data
5. gjøre oppslagsord, artikkeltekst og geografiske kilder søkbare hver for seg og der en kan bruke moderne rettskriving som søkeargument
6. gi alle tekstene en TEI P5 (TEI 2015) mark-up i tråd med moderne edisjonsfilologiske prinsipper

Det er viktig å tenke gjennom de faglig-prinsipielle sidene ved et slikt prosjekt før og under arbeidet. Oddrun Grønvik har diskutert forhold rundt normeringsproblemene i Grønvik (2015). I denne artikkelen vil jeg fokusere på edisjonsfilologi, bruk av gamle ordbokstekster og formidling i en digital verden. Eksempelene er hentet fra Aasen (1873).

3. Det digitale tekstgrunnlaget

Da vi begynte prosjektet, fantes grunnteksten i Aasens ordbøker (1850 og 1873) som elektronisk brødtekst. I forbindelse med Ivar Aasen-jubileet i 1996 var tekstene blitt tastet inn og korrekturlest ved Språksamlingene, Institutt for nordistikk og litteraturvitenskap ved Universitetet i Oslo.

Disse tekstene ble siden brukt som grunnlag for trykte utgaver ved K. Kruken og T. Aarset (blandt andre Aasen 2003). Kruken og Aarset har flettet inn tilleggene i hovedteksten og gjort en lett omredigering som dessverre ikke er markert i teksten.

Hans Ross' ordbok med tilleggene ble også tastet inn og korrekturlest ved Språksamlingene i forbindelse med den storstilte digitalisering av kilde-materialet til Norsk Ordbok på 1990-tallet.

Nettstedet bokhylla.no er lagd av Det norske nasjonalbiblioteket og inneholder i dag faksimiler av om lag 500.000 bøker inkludert verkene til Aasen og Ross. Tekstene er OCR-lest, men resultatet er ikke korrekturlest. For bøker satt i fraktur som er tilfellet for Aasen og Ross' verker, er resultatet svært dårlig.

De elektroniske tekstene fra 1990-tallet er basert på førsteutgavene og er komplette, men er typografisk forenklet. I den inntastete teksten er fraktur erstattet med antikva, liten rettskrift med kursiv og uthevet fraktur med halvfet. Antikva i originalen er ikke markert i avskriften. De er det beste utgangspunktet for vårt edisjonsfilologisk orienterte prosjekt, men vi har likevel valgt å legge inn pekere til faksimilene på bokhylla.no slik at det er mulig å konsultere det originale skriftbildet.

4. Leksikalsk database?

For 20–35 år siden hadde man, spesielt innenfor datalingvistiske miljøer, en viss forhåpning om at en med utgangspunkt i eksisterende ordbøker kunne lage leksikalske databaser for datalingvistiske anvendelser. Den grunnleggende tanken var at den grammatikalske informasjonen og ikke minst betydningsinndelingene i ordbøkene representerte en objektiv beskrivelse av det aktuelle språket, (Kilgariff 1993). Dette var ikke en urimelig antakelse, siden mange ordbøker har vært og fortsatt betraktes som en autoritativ beskrivelse av et språk, fordi ordbøker er ordnet i artikkelform med overskrift og en del faste kategorier og likner på en databaseutskrift, og ikke minst fordi en ordbok representerer en svært stor investering i form av arbeidstid.

Men som leksikografen Sue Atkins har bemerket¹, ville det være en bedre ressursanvendelse om en ga en gruppe leksikografer i oppgave å lage en leksikalsk database fra bunnen av enn å la en gruppe datalingvister prøve å omforme en ordbok til en leksikalsk database. Ingen pre-digital ordbok er systematisk nok eller har klare nok grenser mellom tekstkategoriene til å være egnet til det. Gode og forståelige beskrivelser av ords betydning

¹ I lunsjsamtale under *CompuLex'92*, Budapest 1992.

beregnet på å leses av mennesker, vil sjelden være analytiske nok, jf. Hanks (1979).

5. Dokumentasjon av kulturarv og etterrettelighet

Språk og navn er ikke-materielle fenomener og hører til vår immaterielle kultur(arv) slik som Det norske kulturdepartementet også påpeker:

UNESCO-konvensjonen om den immaterielle kulturarven frå 2003 har mellom anna til formål å verna språk som uttrykksmiddel for denne arven, til dømes språk i form av munnlege tradisjonar og uttrykk. Og i konvensjonen om å verna og fremja eit mangfald av kulturuttrykk [...] er det i dei innleiande merknadene streka under at språkleg mangfald er ein vesentleg del av det kulturelle mangfaldet (St.m.35 2007).

Noen vil kunne innvende at det er farlig å sette språk i bås med kulturarv fordi omverdenen lett vil kunne se leksikografi som en støvete museal og ikke-vitenskapelig aktivitet. Det er selvfølgelig tøv. For det første er samtidskulturen av kort varighet og blir i morgen kulturarv fra i går. Enhver beskrivelse av et språk må være retrospektiv. For det andre er vitenskapelighet knyttet til metode- og etterprøvnbarhet av resultatene. Men også i en vitenskapelig leksikalsk beskrivelse av et språk vil tolkningen av ordenes mening være preget av den kulturelle konteksten forskeren lever i. Ordbøker blir med tiden en historisk dokumentasjon av de kulturelle syn som hersket på tiden de ble skrevet, jf. definisjonen av hottentott som ”innfødt tilhørende en lavtstående syd-afrikansk folkerase” i Norsk Riksmålsordbok (NRO). Eldre ordbøker kan også gi detaljert informasjon om foreldete praktiske forhold som ikke kommer med i ”samtdsordbøker”.

Dette bringer oss inn på et annet aspekt, nemlig ombruk av gamle ordbøker. I forlagsbransjen har det ut fra rasjonalitetsbetraktninger ikke vært uvanlig å foreta en enkel revisjon av eldre ordbøker og presentere dem som nye og oppdaterte. Dette er en praksis som blir mer og mer diskutabel desto lengre tid det har gått siden den originale ordboken utkom. En eldre ordbok er en tekst som i sin tid ble skrevet for å bli lest og brukt av et bestemt

publikum. Å revidere og gi ut en gammel ordbok uten tydelig å klargjøre hva som er nytt og gammel er ikke etterrettelig.

Ordboksredaktørene på 1800-tallet organiserte ikke sine manuskripter etter databasemodeller eller XML-skjema. Men mange eldre ordbøker er skrevet med en imponerende konsistent artikkelstruktur. Både Ivar Aasen og Johan Fritzner (Fritzner 1883) har en veldig ryddig artikkelstruktur i sine ordbøker. Hos Hans Ross er strukturen i artiklene temmelig fri og mye løsere enn hos Aasen. Men i Aasens ordbok fra 1873 ser en også eksempler på at den er skrevet for å leses som en sammenhengende tekst. Se for eksempel 'and' (metafysisk begrep) og 'and' (fuglen) i Aasen (1873). En moderne ordboksdatabase svarer til å klippe opp ordboksteksten til en bunke løse papirlapper. På 1800-tallet falt det neppe noen inn å publisere en ordbok på den måten. Aasen så teksten som en helhet.

6. Ivar Aasens ordbok

Hvor har vi tenkt å plassere arbeidet med Aasens ordbok i dette bildet? Det er et mål for oss å følge moderne edisjonsfilologiske prinsipper. En kan grovt dele edisjonsfilologien inn i den tradisjonelle, der målet er å lage en beste tekst på grunnlag av tilgjengelige tekstvitner, og nyfilologien (new philology, se f.eks. Nichols (1990)) der en ser hvert tekstvitne (manuskript) som et studieobjekt i seg selv. Vårt prosjekt heller i nyfilologisk retning: I den transkriberte teksten skilles det riktignok ikke mellom alle typesnittene i originalen. Linjeskift og spalteskift er ikke markert. Det er derfor ikke mulig å vise den elektroniske teksten i en "faksimilær" gjengivelse. Men dette kompenseres med tilgang til faksimiler av originalen. I den foreløpige digitale utgaven av Aasens ordbok fra 1873 er det bare hovedteksten som gjengis. Aasens egne tillegg og tilleggene til Hans Ross er ikke med. I den fullstendige digitale utgaven vil alle ordbøkene bli utgitt separat i den nyfilologiske tradisjonen.

En kan fundere på om Ross noen gang forestilte seg muligheten av å utgi en samredigert utgave av Aasens ordbok og hans egne tillegg. Det er mulig at han hadde for stor respekt for Aasen til å våge det, at Aasen motsatte seg det eller at det ikke fantes ressurser. I en videreutviklet versjon vil vi eks-

perimentere med en felles utgave basert på hyperlenker og eventuelt sam-sortert visning av de ulike ordbøkene med tilleggene.

7. Tolkning av teksten og indeksering av ord og navn

Som nevnt i innledningen, er formålet å bevare Aasens tekst mest mulig ufortolket og la vår egen tolkning representeres gjennom ekstra TEI-koding av teksten. Et annet mål er å trekke ut informasjon om hva vi mener Aasen anså som gangbare ordformer i ordinventaret i norsk og, om mulig, hvor i Norge Aasen mente de ble brukt. Ordformene eller oppslagsordene vil deretter bli koplet til resten av dokumentasjonsdatabasen. Denne koplingen vil gå gjennom den felles ordindeksen, Metaordboka, og dermed også til selve *Norsk Ordbok*.

skorra, v.n. (ar), glide paa Iis (= skreida, renna seg). Smaal. Nt. schurren. (Schambach 188). Ved Mandal **sko-renna**: glide med bare Sko, uden Skøiter.

```
<entry type="hel" n="26232 1" xml:id="A2_orig27065"><form
n="A2_orig33271" type="origlemma">skorra</orth>,
<gramGrp><pos>v.n.</pos></gramGrp> <form n="A2_orig33272"
type="inflected_"> <orth>(ar)</orth></form></form>, <sense><def> glide paa
Iis (= skreida, renna seg). <placeName xml:id="p_026157" n="5,
Austf">Smaal.</placeName> <rs type="lang">Nt.</rs> schurren.
(<ref><bibl><title>Schambach</title> 188</bibl></ref>). Ved <placeName
xml:id="p_026158" n="323, A-VAgd">Mandal</placeName> <hi
rend="bold">sko-renna</hi>: glide med bare Sko, uden Skøiter.
</def></sense></entry>
```

Figur 1: Artikkelen for oppslagsordet *skorra* vist som tekst og vist med full xml-oppmerking.

En slik kopling vil kunne gi leseren et inntrykk av at Aasen regnet ordformen som en del av hans utkast til det norske skriftspråket. Men egentlig er det vår tolkning av teksten som ligger til grunn. Å finne årsaken til at et stedsnavn er referert i en artikkel, er adskillig mer problematisk. Angir stedsnavnet maksimalutbredelse av et oppslagsord eller at Aasen har

observert en bestemt dialektrealisering av ordet der? Det er viktig å ha en gjennomtenkt forståelse av hva som skal tagges.

Figur 1 viser et eksempel på tagging av en ordartikkel. Vi kan regne med at hovedoppslagsord står på sin vanlige plass først i en artikkel. Videre kan vi regne med at ordformer med ordklassemarkering i gitte omgivelser kan sees som et oppslagsord. Det er viktig siden vi baserer oss på et halv-automatisk utplukk av oppslagsordkandidater. Nærlesing av teksten viser at ikke alle uthevede ordformer er egne oppslagsord, og det finnes ord som ikke er uthevet men som likevel kan regne som oppslagsord. I mange tilfeller kan en bruke signalord og signalfraser til å lete dem opp. Ordet ‘hertil’ indikerer gjerne at neste ord er enten en sammensetning der hovedoppslagsordet i artikkelen inngår som ledd eller at det kommer en avledning av hovedoppslagsordet. I eksempelet i Figur 2 er det derfor rimelig å tolke ordet *tjeldvev* i artikkelen for *tjeld* som et selvstendig oppslagsord. Aasen bruker ‘I Sammensætning’ som innledning til en liten forklaring om hvilken form hovedoppslagsordet har som ledd i sammensetninger for så å ramse opp noen eksempler på slike, se artikkelen for *Kvitel* i Figur 3. Ingen av sammensetningene som nevnes der har egne artikler i ordboken, og de er vel av Aasen betraktet som vanlige, transparente sammensetninger som han har observert. At den første er ført med ordklasse og definisjon, er vel mest for å skille den fra *garn* i betydningen nett.

Tjeld, n. 1) Telt, Hytte af Seildug osv. Nordl. i Formen **Kjeld** (Kjell), især om et Skjul af udspændte Seil. (Andre St. kaldet “Telt”, som er den nedertydske Form). G.N. *tjald*; Sv. tjäll. – 2) et Forhæng, et hængende Tæppe. Hall. (Kjeld). Ogsaa: et Tørklæde, en Dug at kaste over Skuldrene. Rbg. – 3) et Sengetæppe (= Kvitel), Uldtæppe at ligge paa (eller egentl. at indsvøbe sig i). Sæt. Tel. Num. (Kjeld). Hertil **Tjeldvev** (Kjeldvæv), m. Tæppevæv.

Figur 2: Artikkelen for oppslagsordet *tjeld*.

Formen *Kjeld* i Figur 2 er uthevet av Aasen. Det finnes en egen artikkel for *Kjeld* med henvisning til *Tjeld* i betydningen telt eller teppe. Så her er det en kryssreferanse. Formen *sko-renna* i Figur 1 er også uthevet av Aasen, men her finnes det ikke en egen artikkel. Utplukksprogrammet er forsiktig og har ikke markert noen av disse formene som oppslagsformer. Formen *sko-renna*

er eget oppslagsord i *Norsk Ordbok*. Det skyldes vel at det finnes andre belegg for det i seddeldatabasen til NO2014.

I artikkelen for *kvitel* (i Figur 3) er *Gvitel* uthevet av Aasen. Men denne formen er ikke et oppslagsord, men en avvikende målform fra Numedal som Aasen mener det er verd å nevne. I samme artikkel står det 'Ellers kaldet Blæja og Tjeld'. Dette kunne indikere at det er to dialektformer Aasen ikke betrakter som oppslagsordkandidater. Men begge har egne artikler i ordboken. Konklusjonen er at et slikt utplukksarbeid med hell kan gjøres ved hjelp av en utplukksalgoritme, men tvilstilfellene må kontrolleres av en kompetent leksikograf. Oddrun Grønvik (Grønvik 2015) har beskrevet dette i mer detalj.

Kvitel (i'), m. Sengeteppe, Uldteppe at ligge paa. B. Stift, Vald. Hall. Tel. og flere. Aftvig. **Gvitel**, Num. (Jf. Brotkvitel, Mjukkvitel). G.N. *hvitill*. Eng. Dial. *whittle*. Ellers kaldet Blæja og Tjeld. Fleertal **Kvitlar** (nogle Steder Kvislar). I Sammensætning tildeels Kvitte (Kvitla), som **Kvitlegarn**, n. (Traad til Sengetæpper), **Kvittevarp**, **Kvitteveft** og flere.

Figur 3: Artikkelen for oppslagsordet *kvitel*.

Rene kryssreferanser er ganske greie å identifisere siden en her har "fasiten" i form av den komplette oppslagsordlisten. En kan også plukke ut referanser til ord i andre språk og til litteratur som Aasen viser til i ordboksteksten.

Det er mange stedsangivelser i Aasens ordbok, om lag 37 000. Men det er langt fra slik at en kan regne med at stedsangivelsene gir noen oversikt over ordenes utbredelse. Alminnelige ord med utbredelse over hele Norge er ikke geografisk markert. Det synes mer å være spesielle betydninger eller spesielle former som er angitt med sted. Stedsangivelsene i Aasen kan ikke brukes til å lage utbredelseskart for enkeltord. Vi har likevel tatt oss bryet med å tagge stedsnavnene. De er så knyttet til herredsinndelingen i Norge i 1947 som er det året antall herreder var på sitt høyeste, noe over 700. Denne geografiske inndelingen av Norge brukes også i *Norsk Ordbok*. Herredene følger topografien og ferdselsårer. Det er sjelden at vesentlige dialektfenomener deler et herred i to. En kan derfor bruke herredene som geografiske byggeklosser. En del av stedsbetegnelsene hos Aasen refererer til fogderier. Et fogderi var delt i herreder som senere kan ha blitt delt i mindre herreder inntil denne inndelingen nådde sitt maksimum i 1947. Ved å bruke

1947-inndelingen kan en med stor nøyaktighet tilordne et eller flere herreder til stedsnavnene i Aasen (1873). I den elektroniske utgaven er det mulig å søke både på originale stedsnavn, herredsnavn fra 1947, fylker og landskap. Som en hjelp for leseren er det også lagt inn ”mouseover” slik at det moderne navnet eller en liste av slike dukker opp.

8. Hva skal vi være forsiktige med?

En kan ikke alltid være sikker på hvorfor Aasen har ført opp en dialektform og i hvilken relasjon den står til resten av artikkelteksten. Det er heller ikke alltid mulig å vite hva et stedsnavn indikerer. Det kan være vanskelig å gi artiklene en streng formell struktur slik vi gjerne gjør det i et moderne XML-basert redigeringsystem. Likevel kan en indre analyse være formålstjenlig om man vil vise større artikler på en mer pedagogisk måte eller man vil prøve å flette inn Ross’ tillegg i Aasens ordbok. I den grad vi skal prøve å gjøre det, må resultatet av en slik strukturanalyse alltid sees som en subjektiv analyse og markeres som en annotasjon til den originale ordboksteksten.

Det er den originale ordboksartikkelen som er den grunnleggende enheten og som må stå i sentrum. All koding og alt uttrekk til lister over oppslagsord og stedsnavn er hjelpemidler for å få bedre tilgang til informasjonen i ordboken. Dette strenge synpunktet utelukker ikke at vi kan bruke informasjonen i ordboken i et større sammenlenket informasjonssystem som for eksempel databasen (Metaordboka) over kildematerialet i NO2014-prosjektet. Vi må bare være nøye med å oppgi hvilken tolkning av teksten informasjonen er basert på. Dette er normal vitenskapelig kildekritikk. Men i en digital kontekst og spesielt på internett er det ikke alltid så lett å skille mellom originalinformasjon og avledet informasjon. Det er heller ikke noen sterk tradisjon for å være nøye med det på nettet.

9. Hvordan tilrettelegge ordboken til digital utgave?

I en digital utgave av en ordbok, gammel eller ny, er det vanlig at grensesnittet gir brukeren anledning til å søke blant oppslagsord, men noen ganger også foreta fritekstsøk i artikkeltekstene. Man får da gjerne opp en resultatliste med utsnitt av de aktuelle artiklene og eventuelt også en

alfabetisk liste over oppslagsord før og etter den artikkelen som en ønsker å se i sin helhet. Et slikt grensesnitt betrakter en ordbok som en samling av uavhengige og frittstående artikler. Men for eldre ordbøker, som oftest er forfattet som en sammenhengende tekst med kryssreferanser av typen ”se også nedenfor”, bør en ha en mulighet til å rulle forover og bakover i den retro-digitaliserte ordboksteksten og se flere artikler på en gang. Det bør også være mulig å få se faksimiler av den trykte originalen. Det siste er vesentlig for etterretteligheten til digitale utgaver, selv om den digitale utgiveren har brukt mye tid på å gjenskape originalen som digital tekst.

Aasen (1873) har en norsk ortografi som skiller seg vesentlig fra den moderne normen. I tillegg er forklaringen skrevet på et alderdommelig dansk. Stedsnavnene er ikke alltid like forståelig for dagens lesere. Det har derfor vært viktig å kople oppslagsordene til moderne former. Det er gjort gjennom den såkalte Metaordboka som også har manuskriptdatabasen til NO2014-prosjektet og flere moderne ordbøker koplet til seg.

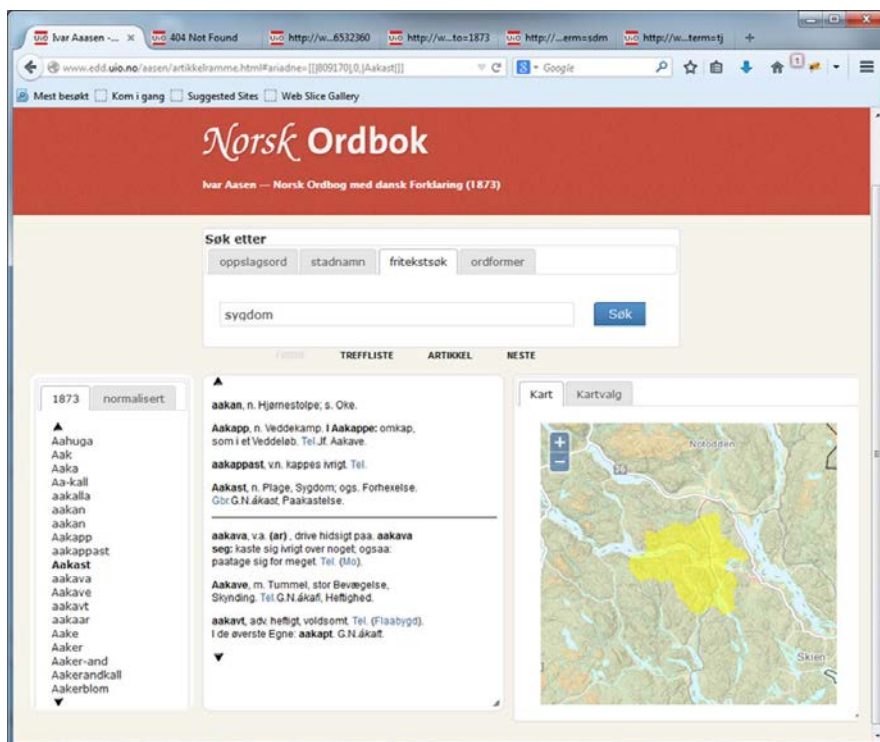
Stedsnavnene er som tidligere forklart, lenket til de norske herredene fra 1947. Nå er ikke det tale om helt moderne enheter, men navnene burde være forståelige. Siden herredsgrensene fra 1947 er digitaliserte og georefererte, kan vi bruke kart både til å vise hvilke(t) område i Norge hvert 1947-herred svarer til og bruke et digitalt kart som søkeinnngang.

I Norge er kunnskapen om 1800-tallsdansk for nedadgående. Som en hjelp kan det derfor være en idé å få tilgang til en fullformsliste over danske ordformer fra 1800-tallet og på den måten muliggjøre enkle oppslag i ODS som kan trigges ved å holde markøren over et ord og for eksempel høyreklikke (eller tilsvarende på en telefon). I ODS er det mange referanser til norske dialektord. Om disse var markert, ville en fra digitale utgaver av Aasens ordbøker eller *Norsk Ordbok* (NO2014) kunne hente dem opp. Tilsvarende kunne en tenke seg at en fra ODS kunne få tilgang til de norske ordbøkene.

I stedet for å lage slike skreddersydde koplingsløsninger for hvert ordbokspar, kunne det være verd å tenke gjennom en litt mer generalisert løsning. De fleste ordbøkene, i alle fall i Norden, inneholder et (varierende) utvalg av de samme informasjonskategorier som oppslagsord, ordklasse, stedsangivelser og litteraturhenvisninger, for å nevne noen. Om en kunne enes om et felles sett av slike kategorier, så kunne en gjennom såkalt ’Linked Data’-grensesnitt åpne for kryss-søk og uthenting av på tvers av

ordbøkene. Det foreløpige grensesnittet for Aasens ordbok er basert på bare tre enkle grensesnitt eller kommandoer for søk etter artikler, stedsnavn eller oppslagsformer. De kan brukes fra andre digitale ordbøker.

- Artikkelsøk: www.edd.uio.no/aasen/a2/artikkel?type=opp&soek=tj*
- Navneliste: www.edd.uio.no/aasen/a2/stadliste?orto=1873&term=s*
- Formliste: www.edd.uio.no/aasen/a2/formliste?orto=1873&term=t*



Figur 4: Enkelt grensesnitt mot den digitale utgaven av Aasens ordbok.

10. Oppsummering

I denne artikkelen har jeg diskutert digital nytgivelse av eldre ordbøker og en del prinsipielle forhold som en må ta i betraktning om digitaliseringen skal yte originalen rettferdigheten og samtidig framstå som både lettbrukt og

etterrettelig. Kort oppsummert: Ved digital nyutgivelse bør en anvende edisjonsfilologiske metoder. En ordbok er et verk som speiler den tiden den ble skrevet i. Endringer i den opprinnelige ordboksteksten representerer nytgiverens subjektive tolkninger og må markeres tydelig. Slike endringer bør sees som annotasjoner og helst legges i teksten som XML-koding eller som 'stand-off' koding i separate filer. En bør tilby faksimiler av originalen. Den digitale utgaven bør ha et åpent grensesnitt slik at andre kan lenke til med såkalte 'Linked Data'-mekanismer. Da kan den digitale ordboken inngå i et større nettverk med de fordeler det gir.

Litteratur

Ordbøker

- Aasen, Ivar (1965): *Norsk Grammatik*. Omarb. Udg. af "Det Norske folkesprogs Grammatik", 3die uforandr. Udg. (Opph. utg. Christiania 1864). Oslo: Universitetsforlaget.
- Aasen, Ivar (1918): *Norsk Ordbog med dansk Forklaring*. Fjerde uforandrede udgave. Kristiania: Vestmannalaget & Cammermeyers forlag.
- Aasen, Ivar (2003): *Norsk Ordbog med dansk Forklaring*. Utg. ved Kristoffer Kruken og Terje Aarset. Oslo: Samlaget
- Fritzner, Johan (1886-1896): *Ordbog over det gamle norske sprog*. Omarbeidet, forøget og forbedret udg. Kristiania: Den norske Forlagsforening.
- GMANUS = *Grunnmanuskriptet* (1935) < www.norskordbok.uio.no >.
- Metaordboka <www.norskordbok.uio.no>.
- NO = *Norsk Ordbok I-XI* (1950-2013). Oslo: Det norske samlaget.
- NRO = *Norsk Riksmålsordbok I*, annet halvbind (1937). Utarbeidet av Trygve Knudsen og Alf Sommerfelt. Oslo: Aschehoug.
- Ross, Hans (1971). *Norsk ordbog, tillæg til "Norsk ordbog" af Ivar Aasen*. Uforandret opptrykk av utg. 1895 + senere tilleg. Oslo: Universitetsforlaget.
- Ross, Hans (1895-1912): *Tillæg 1-6 til "Norsk ordbog" af Ivar Aasen*. Kristiania: Alb. Cammermeyers Forlag.

Annen litteratur

- Grønvik, Oddrun (2015): Eit lemmaregister for Ivar Aasens Norsk Ordbok med dansk forklaring i databaseformat. I: Caroline Sandström m.fl. (red): *Perspektiv på leksikografi, grammatik och språkpolitik i Norden*. Helsingfors: Institutet för de inhemska språken och författarna, 81-99.
- Hanks, Patrick (1979): To what extent does a dictionary definition define? I: R. R. K. Hartmann (ed.): *Dictionaries and their Users, papers from the 1978 B.A.A.L. Seminar on Lexicography*. Exeter: University of Exeter, 32-38.
- Haugen, Einar (1972): Analysis of a sound group: SL and Tl in Norwegian. I: Evelyn S. Firchow (red.): *Studies by Einar Haugen, Presented on the Occasion of his 65th Birthday, April 19, 1971*. Berlin & The Hague: de Gruyter, 115-141.
- Kilgariff, Adam (1993): Dictionary Word Sense Distinctions: An enquiry into their nature. I: *Computers and the Humanities* 26 (1-2), 365-387.
- Nichols, Stephen G. (1990): Introduction: Philology in a Manuscript Culture. I: *Speculum* 65, 1-10. doi:10.2307/2864468.
- NO2014 = Prosjektet Norsk Ordbok 2014. <www.norskordbok.uio.no>.
- St.m.35 (2007) = Stortingsmelding 35 (2007-2008) *Mål og mening*.
- TEI (2015) = Text Encoding Initiative, *Guidelines to Text Encoding (P5)*. <www.tei-c.org>.
- UNESCO (2003) = *The Convention for the Safeguarding of the Intangible Cultural Heritage*. UNESCO, MISC/2003/CLT/CH/14, Paris, 2003. <www.unesco.org/culture/ich/en/convention>

Christian-Emil Smith Ore
førsteamanuensis, dr.scient.
Universitetet i Oslo
Institutt for lingvistiske og nordiske studier
Enhet for digital dokumentasjon
Boks 1102, Blindern
NO-0317 Oslo
c.e.s.ore@iln.uio.no