

Bokmål og nynorsk samindeksert – Metaordboka som verktøy for jamføring og utforsking av ordtilfang

Oddrun Grønvik & Christian-Emil Smith Ore

Two written standards of different origins are in use in Norway, Bokmål and Nynorsk. During the 20th century Bokmål and Nynorsk came closer to each other through a series of orthographic reforms. Results of this policy can only be studied in detail if the standards are linked through the lexical item. A suitable tool is the Meta dictionary, an electronic registry for the Norwegian lexicon which allows linking lexicographical evidence to several headword forms through one node, the Meta Dictionary entry. This paper discusses principles and methods for expanding coverage of the linked standards through adding new, machine-analysed materials.

1. Innledning

Bokmål og nynorsk gir som skriftmål hver sin innfallsport til norsk språkbruk og skriftkultur. Enten en ser bokmål og nynorsk som to ulike skriftmål eller som to ulike skriftlige uttrykk for norsk talemål, så er de nærstående skriftspråkstandarder, og utgjør til sammen den skriftspråklige rammen som folk i Norge lever innenfor, med et historisk spenn fra dansk (før 1907) til Aasen-normalen (uttrykt i oppslagsordformen i Aasens *Norsk Ordbog* fra 1873).

Formålet med arbeidet som er beskrevet i denne artikkelen, er å legge til rette for en systematisk undersøkelse av skriftmålsutviklingen i norsk på 1900-tallet. Dersom en skal kunne uttale seg med en rimelig grad av sikkerhet om resultatene av norsk språkpolitikk etter 1900, må det foreligge språksamlinger som samordner både primærkilder og sekundærkilder for begge målformer på ordnivå.

2. Premisser

Inngangspunktet til vitenskapelig analyse av språk er alltid ord i kontekst, enten tilfanget er registrert som lyd eller tekst. Dette gjelder hele spekteret av lingvistiske deldisipliner, fra fonologi til stilistikk. Det registrerte tilfanget som skal undersøkes må være lett tilgjengelig, ordnet etter faste kriterier og ha adekvate metadata, med kilde og annen bakgrunnsinformasjon på plass.

Formålet er å legge til rette for synkron og diakron jamføring av rettskrivningen for bokmål og nynorsk. Det eneste funksjonelle er en samordning der den leksikalske gjenstanden («lexical item», Atkins & Rundell 2008:163 ff.) er fikspunkt,

representert ved grunnformen, dvs. den eller de ordformene som brukes som oppslagsform for artikler i ordbøker for morsmålsbrukere (jf Bergenholz et al. 1997, s. 132). Forutsetningen for en slik samordning er at egenskapene til den leksikalske gjenstanden er kjent, og at det er enighet om rangering av egenskapene.

Nedenfor er det antatt at en leksikalsk gjenstand har et tilhørende begrep eller et begrepssett som kan defineres med støtte i dokumentert språkbruk, og at bruken kan vises med brukseksempler. Det at det eksisterer dokumentert enighet om betydning og bruk, er nøkkelegenskapen til en leksikalsk gjenstand. En ordform som bare har formopplysninger, men som ikke har eller kan knyttes til en definisjon, kan ikke regnes som en leksikalsk gjenstand, fordi ordets betydning og bruk ikke er kjent.

I tillegg er det forutsatt at ordets formelle egenskaper er enten kjent eller mulige å avlede fra det en vet om språket for øvrig. De formelle egenskapene omfatter ordtype, ordklasse, bøyning, uttale og historie. Det skriftlige uttrykket for en og samme leksikalske gjenstand kan derimot ha ulik form etter kontekst. Rettskrivning i egentlig forstand – ortografi – kan bare eksistere innenfor rammen av et standardspråk. Men spennet av formuttrykk for leksikalske gjenstander omfatter alle former av ordet eller uttrykket som kan finnes i tolkbar tekst eller tale.

De mest systematiske ordbeskrivelsene finnes i større ordbøker der ordboksartiklene for leksikalske gjenstander omfatter alle egenskapene som er nevnt ovenfor.

3. Verktøy

Jamføring av rettskrivningene krever et felles register med maksimum ett sammenkoplingspunkt per leksikalsk gjenstand. I dette registeret må hver leksikalske gjenstand kunne ha både flere normerte grunnformer som oppslagsord, og mange ulike typer brukseksempler med hver sine metadata.

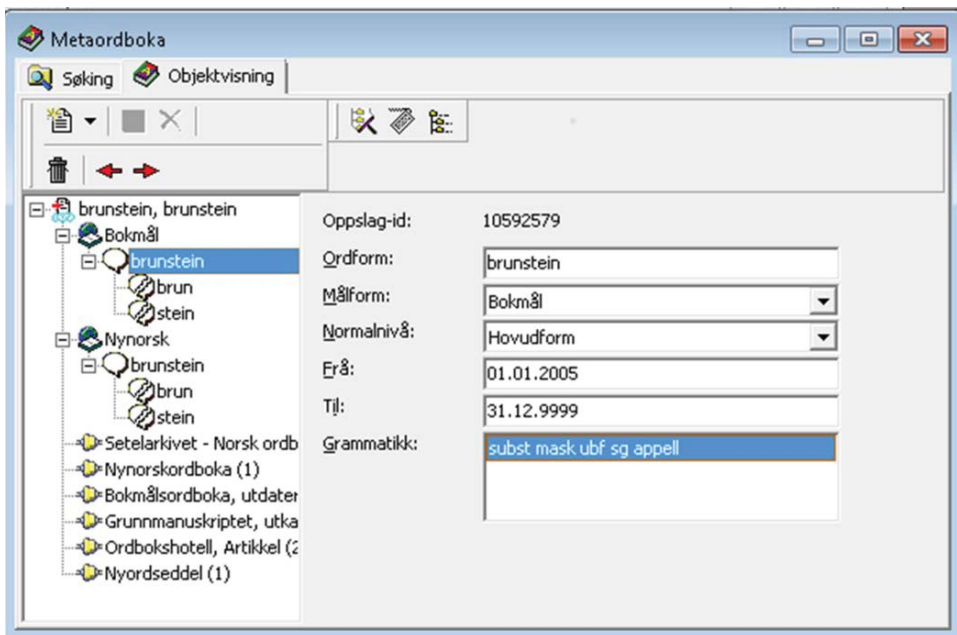
Tidligere brukte man kortarkiv der ordningsprinsippet ville vært den abstrakte størrelsen «den leksikalske gjenstanden med definisjon og brukseksempler», uttrykt ved en eneste foretrukket oppslagsform. Papirutgaven av Setelarkivet til Norsk Ordbok er et slikt arkiv. Oppslagsformen er satt til «Nynorsk, 1938-rettskrivningen, tradisjonelle former». Denne rettskrivningen var eneste mulige inngang til materialet i papirform dersom en ikke lagde ekstra kortarkiv som register til hovedarkivet. Men i den digitale verdenen er det mulig å lage mange innganger til et materiale.

For norsk språk eksisterer det i dag to verktøy som kan brukes som sammenkoplingsregister, Metaordboka og Norsk ordbank. Norsk ordbank er et fullform-system for bokmål og nynorsk. Det finnes to ordbanker, en for hver skriftstandard. De to ordbankene er ikke koplet sammen, men fungerer parallelt. Det andre verktøyet er Metaordboka, som har den funksjonaliteten en trenger for å lage et jamføringsregister for rettskrivningene, og som er brukt i arbeidet som er beskrevet i denne artikkelen (se også Ore 2000). Begge verktøyene er del av språksamlingene ved Universitetet i Bergen.

4. Metaordboka

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. Det er dermed mulig å bygge inn et historisk formregister per artikkel ved å kople til leksikalske ressurser for begge målformer, og indeksere dem. Hver artikkel kan ha ett eller flere oppslagsord i grunnform. Ordformer som er eller har vært normerte former i rettskrivningen, er brukt som oppslagsformer, mens kildene speiler den variasjonen som foreligger i skrift og tale over tid. Opplysningsskjemaet for oppslagsord i Metaordboka gjør at den ligger tett opp til fullformsystemet Norsk ordbank, men Metaordboka har ikke bøyingsinformasjon og regler for å generere fulle bøyningsskjema.

Artikkelkroppen består av referanser eller «hyperlenker» til ulike kildedatabaser. De fleste viser brukseksempler eller gir opplysninger om betydning, men det finnes også belegg som bare gir opplysning om forekomst, uttale, ordklasse, bøyning, ordlagingspotensiale eller liknende. I sin enkleste versjon er Metaordboka en sortering eller gruppering av tradisjonelle ordbokssedler. En metaordboksartikkel svarer da til en konvolutt eller eske med sedler som ansees å behandle samme oppslagsord.



Figur 1: Eksempel på metaordboksartikkelen for «brunstein».

Metaordboka samordner normert og unormert tilfang, tekst og transkribert tale. Materialet som er lenket sammen gjennom Metaordboka kan grovt kategoriseres som primærkilder og sekundærkilder.

Primærkildene omfatter (1) ordsedler, med metadataene oppslagsord, ordklasse, stedfesting eller litterær kilde og medarbeidersignatur. Belegg kan være tekst, lyd, bilde, video, (2) korpuskonkordanser.

Sekundærkildene omfatter utvalgte ordbøker, for eksempel grunn-manuskriptet til *Norsk Ordbok*, *Nynorsk etymologisk ordbok* (Torp 1919), og tidligere utgave av *Nynorskordboka* (NOB) og *Bokmålsordboka* (BOB). I tillegg kommer innholdet i Ordbokshotellet, som er beskrevet nedenfor.

5. Ordbokshotellet

«Ordbokshotellet» er både en søke- og framvisningsportal og et bibliotek for digital(isert)e ordbøker (Tvedt et al. 2007). Det inneholder nå 60 store og små dialektordsamlinger, ordbøkene Ivar Aasens ordbok (Aasen 1873) og *Norsk Landbruksordbok* (NLO) og fire rettskrivingsordlister for bokmål. Det utvides fortløpende og er Språksamlingenes hovedlager for digital(isert)e ordbøker. Hver ordbok/liste er en XML-kodet tekst i henhold til TEI P5 Dictionary. Siden graden av standardisering i foreleggene varierer, har hver artikkel ett eller flere normerte oppslagsord som brukes i koplingen til Metaordboka. Normeringspråket for dialektmaterialet er nynorsk.

For hver ordbok/liste er det angitt om den er normert til bokmål eller nynorsk. For koplingen til Metaordboka betyr det at en ordbok markert med nynorsk koples til en metaordboksartikkel som har oppslagsordet med språkform markert som nynorsk, og tilsvarende for bokmål.

6. Tidligere og nåværende bruk av Metaordboka

Metaordboka var i perioden 2002–2015 indeksert bare for nynorsk, etter 1938-normalen med tradisjonelle former. Formålet var å tilrettelegge resursene for bruk i Norsk Ordbok. Inndelingskriteriet er grunnform og ordklasse. En artikkel i Metaordboka ble ikke homografseparert selv om redigering i *Norsk Ordbok* førte til flere ordboksartikler.

I 2012 fikk *Norsk Ordbok* nettutgave. Dermed kom ønsket om å kunne søke i både *Norsk Ordbok* og Metaordboka med moderne rettskriving, helst fra både bokmål og nynorsk. Tiden var inne for å se på Metaordbokas utvidingspotensiale, ettersom den kunne ha flere oppslagsformer per artikkel, med markering av skriftform og gyldighetsperiode. Hver artikkel i *Norsk Ordbok* er dessuten lenket til en metaordboksartikkel som omfatter materialet NO-artikkelen ble redigert ut fra. I dag er *Norsk Ordbok* direkte søkbar bare i den målformen som ble brukt som

redaksjonsspråk. Men om metaordboksmaterialet får inngang både fra bokmål og nynorsk og i gjeldende rettskriving, vil det være langt enklere å tilrettelegge for at *Norsk Ordbok* også blir søkbar med dagens rettskriving (se også Ore 2012).

Første skritt var derfor å kople opp BOB mot Metaordboka på nytt. I forlengelse av dette ble også *Nyordsarkivet* for bokmål koplet opp mot Metaordboka. Oppkoplingen ble gjort halvautomatisk. De samordningsprinsippene som er beskrevet i neste avsnitt, gav grunnlag for en maskinell samkjøring. Bokmålsordene som manglet blant oppslagsordene i Metaordboka fikk nye rene bokmålsartikler i Metaordboka.

Dette førte til at Metaordboka fikk noe over 250 000 artikler med oppslagsform i normert bokmål. Av disse har om lag 70 000 oppslag både på bokmål og nynorsk. Ettersyn viser at denne forsiktige holdningen har gitt en rekke bokmålsartikler som bør slåes sammen med en tilsvarende nynorskartikkel. Men det krever manuell kontroll å sikre at dette blir riktig gjort.

7. Samordningsprinsipper for bokmål og nynorsk

Bokmål og nynorsk er nærstående skriftmålsstandarder, men langt fra identiske. Om en holder seg til samordning av grunnformer på leksikalsk nivå, forekommer det likevel at (1) en leksikalsk enhet er belagt i den ene målformen, men ikke den andre, (2) homografinnelingen er ulik, noen ganger med utgangspunkt i ortografi, andre ganger betydning eller bruksmåter. I den første sammenkoplingen av bokmål og nynorsk mot Metaordboka er det derfor valgt en restriktiv linje: bokmåls- og nynorsk materialet koples til samme MO-artikkel bare dersom det eksisterer en grunnform som er identisk for bokmål og nynorsk, og som er eller har vært gjeldende rettskriving. I kravet om identitet kommer også leddeling inn. Dersom den ene grunnformen er leddelt og den andre ikke, blir det ingen maskinell oppkopling. Tabell 1 viser hvordan systemet er satt opp.

I tabellen omfatter «full identitet» også at (1) bokmålssubstantiv med genus f og m svarer til nynorsk f, og at (2) nynorsk infinitivsending *-e* og *-a* svarer til bokmål infinitivsending *-e*.

I sammenkoplingsprosessen er det ikke tatt hensyn til etymologi – det er antatt at alle grunnformer i rettskrivningen har sine motsvar i nåværende norsk tale. Dermed forekommer det sammenkoplinger som dette:

Nynorsk: *botnlaus*, adj

Bokmål: *botnlaus* el. *botnløs* el. *bunnlaus* el. *bunnløs*, adj

Type	Bokmål	Nynorsk
full identitet	<i>kaste</i> , v	<i>kaste</i> v
identitet i en form	<i>legge</i> , v	<i>legge</i> el. <i>leggje</i> v
identitet i en form	<i>bo</i> el. <i>bu</i> , v	<i>bu</i> v

Type	Bokmål	Nynorsk
identitet i en form	<i>høgraud</i> el <i>høgrød</i> el <i>høyraud</i> el <i>høyrød</i> , adj	<i>høgraud</i> , adj
identitet i en tidligere form	<i>vanskelig</i> , adj	<i>vanskeleg</i> , adj (til 2012 også [<i>vanskelig</i> adj])
identitet i en tidligere form	<i>høst</i> m (til 1981 også [<i>haust</i> m])	<i>haust</i> m

Tabell 1: Kriterier for sammenkopling mellom bokmål og nynorsk

På noen punkter skiller Metaordboka strengere enn i de tilkoplta ordbøkene og ordlistene. I *NOB* og *BOB* forekommer det at ulike avledninger er ført som oppslagsord til samme artikkel. Det ser ut til særlig å gjelde substantiv som er avledet fra verb:

- bokmålsavledninger med *-else* m og *-ing* m
- nynorskavledninger med *-sel*, m el. *-sel*, f og *-sle*, f.

En rimelig tolking er at en i disse ordbøkene mener at det bare er snakk om rettskrivningsvarianter. At Metaordboka har et strengere skille, gjør det mulig å kople sammen identiske oppslagsord på *-ing* eller på *-sel* på en konsistent måte. For eksempel vil vi kunne kople bokmålsordene *ending* og *tilførsel* sammen med nynorskordene *ending* og *tilførsel* uten å plassere koplingene i sammenheng med bokmål *endelse* eller nynorsk *tilførsle*.

Metaordboka var i utgangspunktet en enkel systematisering av språklig grunnlagsmateriale uten homografseparering. Sammenkopling av bokmål og nynorsk tvinger fram bruken av leksikalsk gjenstand som koplingspunkt.

III **bu** verb

1 ha fast tilhald (i kortare eller lengre tid); ha heimen sin på (ein stad); leve i lag

2 finnast, vere (IV)

3 lage til; gjere ferdig

Figur 2: Artikkel for verbet å *bu* i *NOB*

I *BOB* dekker verbet *bo* el. *bu*, v, betydning 1 og 2. Betydning 3 er skilt ut i et eget verb, *bu*. For å få tatt hensyn til denne oppdelingen må metaordboksmaterialet fordeles på to artikler, en for bokmål *bu* el. *bo*, v, nynorsk *bu*, v og en hvor både bokmål og nynorsk har oppslagsform *bu*, v.

8. Oppkopling av skoleordlister for bokmål

Metaordboka kan brukes til å lage en historisk oversikt over ordinventaret i norsk. Som en test på hvordan det kan gjøres, ønsket vi å se på dokumentasjonen av det sentrale ordforrådet i bokmål fra 1938 og utover. Testmaterialet var fire skoleordlister for bokmål. I 1938 og 1959 var det store og omfattende rettskrivningsreformer for både bokmål og nynorsk som omfattet både ortografi og ordforråd. To av ordlistene er fra 1939, en fra 1959 og en fra 1973. Ordlistene ble maskinelt behandlet slik at oppslagsordene med ordklasse ble hentet ut der det var mulig. Deretter ble listene med oppslagsord kjørt mot oppslagsordene i først bokmålsoppslagene i Metaordboka, og dersom det ikke var treff, ble nynorskoppslagene brukt. Omtrent ti prosent av oppslagsordene i ordlistene manglet ordklasse, eller gav flere tilslag i Metaordboka. De ble sjekket manuelt. Deretter ble de fire ordlistene gitt en enkel TEI-koding, lagt inn i Ordbokshotellet som bokmålskilder og koplet opp mot eksisterende artikler, eller lagt i nye artikler. Samordningsprinsippene for bokmål (og nynorsk) er beskrevet over.

De fire ordlistene gav tilsammen ca. 70 000 nye koplinger til Metaordboka, fordelt på ca. 30 000 metaordbokartikler. Drøyt 7 600 oppslagsformer finnes i alle fire ordlister. Noe over 12 600 oppslagsord finnes i bare en av ordlistene. Alle ordlistene har mange ord som ikke forekommer i de andre.

Ordlister	Utgitt	Antall oppslagsord	Ordform uten ordklasse	Antall tilslag i MO-artikkel ved maskinell analyse	MO-artikkel bokmål ikke funnet
Lange	1939	12 422	454	10 279	1 761
Eitrem	1939	16 626	2 561	12 054	2 721
Krogstrup & Seip	1959	16 657	559	14 178	1 903
Torvik	1973	23 535	837	20 153	2 957

Tabell 2: Oversikt over tilkoblede ordlister, automatisk tilslag og artikler som krevde manuell kontroll.

Den minste har om lag 11 prosent «særegne» ord. De to jamstore fra 1939 og 1959 har henholdsvis 18 og 8 prosent. Den største ordlisten, fra 1973, har flest både relativt og absolutt, med 29 prosent ord som ikke forekommer i de andre, men deler like fullt 16 500 ord med en eller flere av de 3 andre, spesielt med ordlisten fra 1959 (Krogstrup & Seip 1959). 85 prosent av oppslagsordene i denne er også med i 1973-ordlisten. For den jamstore ordlisten fra 1939 (Eitrem 1939) er 75 prosent av ordene med i 1973-ordlisten. En rask gjennomgang viser at svært mange av de 25 prosentene fra denne 1939-ordlisten som ikke er med i 1973-ordlisten, faktisk er rettskrivingsnøytrale ord som *lovmessig* og *metafysikk*.

Foreløpige resultater viser at de fleste oppslagsformene som ble introdusert i 1938, finnes i dagens bokmålsnorm. Det kreves mer arbeid for å få et fullt bilde, og det er mange spørsmål i tilknytning til dette arbeidet som fortjener videre utforskning. Eksperimentet med de fire ordlistene viser at det er overkommelig å lenke inn ordlister fra ulike perioder og på den måten gi et historisk overblikk over de leksikalske gjenstandene og rettskrivingsnormene.

Et tilsvarende overblikk over bruken krever tilgang til et historisk korpus. Metaordboka er utstyrt for å lenke til korpus og til fulle konkordanser. Norsk bokmål skilte lag med dansk skriftspråk for alvor i 1907. Et bokmålskorpus med tekster fra 1907 til i dag står høyt på ønskelisten.

9. Konklusjon

Det er her vist at det er både mulig og rimelig å legge til rette for en systematisk undersøkelse av skriftmålsutviklingen i norsk på 1900-tallet, ved å samordne primærkilder og sekundærkilder for begge målformer med den leksikalske gjenstanden som samordningspunkt. Dette forsøket viser ett resultat av slik samordning. Et annet er at arbeidet i de siste tiårene med språksamlingene for nynorsk også er til nytte for dokumentasjon og utforskning av bokmål, og dermed hele den norske skriftmålssituasjonen.

Litteratur

Ordbøker

- Aasen, Ivar (1873) *Norsk Ordbog med dansk Forklaring. Fjerde uforandrede udgave, (1918)*. Kristiania: Vestmannalaget og Cammermeyers forlag.
- Bergenholtz, Henning et al (1997): *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.
- BOB = Wangensteen, Boye (2006): *Bokmålsordboka. Definisjons- og rettskrivingsordbok*. (3. utg.). Oslo: Kunnskapsforlaget.
- Eitrem, Hans (1939): *Rettskrivning 1938. Regler for skoler og privat bruk*. Oslo: Fabritius.
- Krogsrud, Torgeir & Didrik Arup Seip (1959): *Rettskrivningsregler*. Rettskrivningslære og ordliste etter „Ny læreboknormal 1959“. Oslo: Cappelen.
- Lange, August (1939): *Norsk rettskrivningsordliste*. Oslo: Tiden Norsk Forlag.
- NLO = Rommetveit, Magne (1978): *Norsk landbruksordbok I-II*, Oslo: Samlaget.
- NO = Norsk ordbok: ordbok over det norske folkemålet og det nynorske skriftmålet. 1966-2016. Bd. 1-12. Oslo: Det Norske Samlaget.

NOB = Marit Hovdenak, Laurits Killingbergtrø, Arne Lauvhjell, Sigurd Nordlie, Magne Rommetveit & Dagfinn Worren (2006): *Nynorskordboka* (4. utg.). Oslo: Samlaget.

Nyordsarkivet. Bokmål. Språksamlingane, Universitetet i Bergen.

Torp, Alf (1919): *Nynorsk etymologisk ordbok*. Kristiania: Aschehoug.

Torvik, Ingvald (1973): *Ordlister for alle*. Bokmål. Oslo: Universitetsforlaget.

Annen litteratur

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

Ore, Christian-Emil, (2000): Metaordboken – et rammeverk for Norsk Ordbok. I: *Nordiska studier i lexicografi* 5. Göteborg: Nordisk forening for lexicografi, s. 250—270.

Ore, Christian-Emil Smith (2012): Nettordbøker og Norsk Ordbok – hvordan etablere en vitenskapelig nettordbok. I: *Nordiska studier i lexicografi* 11. Rapport från Konferensen om lexicografi i Norden, Lund 24–27 maj 2011. Lund: Nordisk forening for lexicografi, s.486–499.

Tvedt, Lars Jørgen, Elisabeth Lien & Øyvind Eide (2007): Ordbokshotellet – varig lagring og formidling av norske ordsamlinger. I: Torben Arboe (red.), *Nordisk dialektologi og sociolingvistik*, Foredrag på 8. Nordiske Dialektologkonferens, Århus 2006. Peter Skautrup Centret for Jysk Dialektforskning, Aarhus Universitet, s. 379–388.

Oddrun Grønvik
Forsker emeritus
Universitetet i Oslo
Institutt for lingvistiske og nordiske studier
Boks 1102, Blindern
N-0317 Oslo
oddrun.gronvik@iln.uio.no

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