

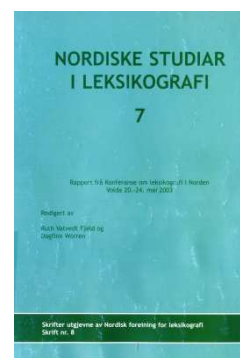
NORDISKE STUDIER I LEKSIKOGRAFI

Titel: Metaordboka som reiskap for lemmaseleksjon og ordboksproporsjonering

Forfatter: Ålov Runde, Terje Svardal, Oddmund Vestenfor

Kilde: Nordiske Studiar i Leksikografi 7, 2005, s. 326-330
Rapport frå Konferanse om leksikografi i Norden, Volda 20.-24. maj 2003

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.

Ålov Runde, Terje Svardal, Oddmund Vestenfor

Metaordboka som reiskap for lemmaseleksjon og ordboksproporsjonering

Metaordboka (The Metadictionary) is a search index to the databases (digitalized archives) of Norsk Ordbok. In addition to the digitalized original paper slip archive a digital corpus of written Norwegian is being developed. This is the new digital platform for the further edition of Norsk Ordbok.

In this paper two components of dictionary making, i.e. lemma selection and dictionary volume planning, are focused upon. The Metadictionary gives an exact survey of the number of lemmas in the database. This, combined with our knowledge about different types and number of sources connected to each lemma gives us information that enables us to set up a plan for the contents of every volume of the dictionary.

Innleiing

I dette innlegget vil vi komme inn på fire hovudpunkt:

- 1) Kva er Metaordboka (MO)?
- 2) Korleis kan Metaordboka vere til hjelp i arbeidet med lemmaseleksjon?
- 3) Korleis kan vi nytte Metaordboka i arbeidet med proporsjoneringa av Norsk Ordbok?
- 4) Korleis kan bruken av korpus supplere (det allereie eksisterande) arkivmaterialet til Norsk Ordbok?

1 Metaordboka

Metaordboka er, enkelt sagt, ein systematisert søkjeinngang til dei ulike elektroniske samlingsane til Norsk Ordbok (Ore 1999, Grønvik 1997). Ho er normert på grunnlag av Skard (1939) som er ei ordliste over rettskrivinga for nynorsk, vedteken i 1938. Vi valde denne norma for Metaordboka, fordi ho er nytta som redaksjonsmål for Norsk Ordbok. I Metaordboka er variantar av lemma førde under eitt og same oppslag.

Metaordboka har no følgjande tilfang: Der finn ein for det første den digitale versjonen av setelarkivet til Norsk Ordbok. Det er ca. 3 millionar setlar som ligg føre som faksimile, og dessutan er opplysningar på setelen om oppslagsord (lemma), ordklasse, kjelde, heimfesting og innsamlar skrivne inn i eit eige skjema. Grunnlaget for søkjeindeksen i MO er oppslagsorda frå setelarkivet.

På same måte er arkivet til Trønderordboka behandla, og desse setlane er også kopla til MO. Dessutan har MO elektroniske utgåver av NOB, Skard 1939 og det såkalla Grunnmanuskriptet. Grunnmanuskriptet er det første utkastet til Norsk Ordbok, som blei ferdigstilt i 1940, men aldri utgjeve.

Ein typisk MO-artikkel kan illustrerast med t.d. artikkelen 'idrott', som vi ser nedanfor

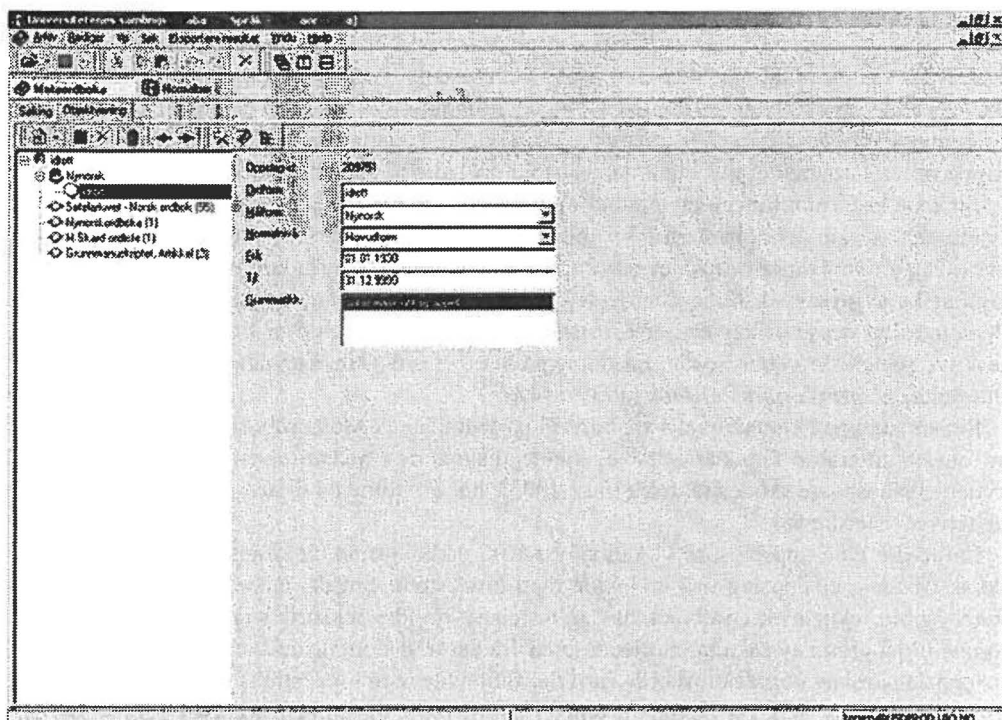


Fig. 1

(Fig. 1). Den er normert til maskulinum på grunn av normeringa i Skard. Til denne MO-artikkelen er det knytt setlar frå arkivet til Norsk Ordbok (55 i alt), og artiklar frå NOB, Skard 1939 og Grunnmanuskriptet.

I samarbeid med Eining for digital dokumentasjon ved HF-fakultetet (EDD, tidl. DOK) arbeider vi kontinuerleg med å bygge ut MO. Mellom anna skal Alf Torp *Nynorsk etymologisk ordbok* (1919) koplast til MO, og på planen for neste år står også å leggje inn elektroniske utgåver av Ivar Aasens ordbøker frå 1850 og 1873.

2 Lemmaseleksjon

Metaordboka er ein nyttig reiskap i arbeidet med Norsk Ordbok, mellom anna når det gjeld spørsmålet om kva for oppslagsord som skal vere med i ordboka. Her skal det kort nemnast nokre sentrale punkt knytte til dette spørsmålet. Kriteria som ligg til grunn for lemmautvalet, er både positivt og negativt definerte.

Metaordboka skal vere grunnlaget for oppretting av artiklar i Norsk Ordbok. Desse skal i første omgang genererast utifrå dei kriteria som blir nemnde nedanfor.

Metaordboka gir oss for det første eit fullstendig oversyn over talet på oppslagsord i databasen. Det ligg no på 554 275 ord. I Norsk Ordbok, som er planlagt til 12 band, skal om lag 300 000 ord beskrivast. Vi har med andre ord eit overskot på om lag 250 000 ord.

Av oppslaga i Metaordboka er ca. 50 prosent hapax legomenon, altså ord som berre har éi kjelde knytt til seg. Her er hovudregelen at ord med berre ei kjelde bak seg ikkje skal beskrivast i Norsk Ordbok, men der finst nokre viktige unntak frå regelen. Målførebelegg skal som ein hovudregel beskrivast i Norsk Ordbok, sjølv om dei er hapax. Artiklar frå Grunmanuskriptet blir også tekne med i Norsk Ordbok, sjølv om dette er einaste belegget vi har på ordet. Grunnen er at Grunmanuskriptet byggjer på Aasen og Ross.

Eit anna kriterium for om eit lemma frå Metaordboka skal få vere med i Norsk Ordbok, er kva litterær kjelde det er henta frå. I samlingane våre finst det ekserpt frå ordbøker laga relativt tidleg i nynorskperioden, der eitt av måla tydeleg har vore å komme med framlegg til (ny)norske avløyarsord. Ikkje alle desse orda har slått igjennom og gått inn i skriftspråket. Dersom ei av desse ordbøkene er einaste belegg på ordet, vil vi velje det bort, for ikkje å beskrive reine «skrivebordsord». Eit døme på dette er ordet 'innkrevjande', som er henta frå Ola Raknes: *Fransk-norsk ordbok* (1939-1942).

Ein annan type lemma vi vel å sjå bort frå på grunnlag av Metaordboka, er samansetningar frå omsett litteratur som berre har éi kjelde. Faksimilen nedanfor, eit ekserpt frå Henrik Rytter: *Shakespeare skodespel I-X* (1922-1933), har eit døme på denne typen samansetning, adjektivet 'isterdiger':

Ein tredje type oppslagsord vi kan sjå bort frå på bakgrunn av kunnskap om kjeldene til Norsk Ordbok, er oppslagsord innleidde med bindestrek. Suffiks er beskrivne i andre ordbøker og blir tekne med i ordboka, medan resten av «bindestreksorda» er oppslagsord som er komne til på grunn av såkalla etterleddssetlar frå det tradisjonelle arkivet. Vi vil berre generere oppslagsord av «bindestreksord» (artiklar innleidde med bindestrek) som byggjer på ordboksartikkel. Dvs. at dei er beskrivne anten i Skard 1939, Grunmanuskriptet eller NOB. Til dømes vil det bli skrivi ein artikkel om suffikset *-sam*.

Til saman gir kombinasjonen Metaordboka og kjennskap til kjelder eit godt grunnlag for lemmaseleksjon. Samstundes er det viktig å understreke at all informasjonen i databasen er intakt og tilgjengeleg for redaktørane i arbeidet med Norsk Ordbok. Det er ikkje slik at vi kvittar oss med informasjon, men vi gjer eit utval av det vi vil vise i ordboka.

3 Ordboksproporsjonering

Av dei 12 banda som NO skal omfatte, er 4 allereie utgjevne¹ og åtte står att å skrivi fram til 2014. Kwart band i Norsk Ordbok omfattar om lag 25 000 ordartiklar.

Som nemnt tidlegare i artikkelen, inneheld arkivet i alt om lag 555 000 MO-artiklar. Bolken h-å, som står att å redigere, omfattar 367 000 MO-artiklar. Deler vi dette talet på to, utfrå kriteriet om at hapax ikkje skal redigerast, er vi nede i 180.000 ord, som gir eit snitt på 23 000 artiklar per attståande band.

For planlegginga sin del kan vi altså bestemme nokså nøyaktig innhaldet i og omfanget av kvart band. Kvar bokstav er kartlagd; bokstaven *s* har til dømes ca. 85 000 ord. Av desse skal ca. 50 prosent beskrivast (dvs. om lag 42.000 ord), som det går fram av utvalskriteria. Dette utgjer ca. 2 heile band. På den andre sida er det på bokstaven *i* knapt 8 000 MO-artiklar og den vil omfatte knapt 4 000 artiklar i Norsk Ordbok. Dette utgjer grovt rekna 1/6 ordboksband.

4 Korpus som supplement

Korpuset vil kanskje særleg vere til hjelp i arbeidet med å finne kollokasjonar og gi betre grunnlag for semantisk inndeling av tilfanget. Dessutan kan ein bruke korpuset til å generere nye oppslag i Metaordboka, der det er manglar i arkivet til Norsk Ordbok.

5 Oppsummering

Metaordboka er ein svært nyttig reiskap i arbeidet med Norsk Ordbok på fleire måtar; her er det nemnt to av dei. Vi har sett korleis ein ved hjelp av Metaordboka, kombinert med kjennskap til kjeldene til Norsk Ordbok, har eit godt grunnlag i arbeidet med lemmaseleksjon. Dessutan kan vi ved hjelp av Metaordboka bestemme nokså nøyaktig innhaldet i og omfanget av kvart band som står att å redigere. Og sist, men ikkje minst: Metaordboka supplert med korpusbruk gir eit rettare og meir fullstendig bilete av det moderne nynorske skriftspråket.

Litteratur

- Grønvik, Oddrun 1996: *Om kjeldegrunnlaget for Norsk Ordbok i Norsk Ordbok – nynorskens leksikografiske kanon?*, Oslo, 23-37
- Ore, Christian-Emil 1999: *Metaordboken – et rammeverk for Norsk Ordbok*, Gøteborg 1966-: Norsk Ordbok – ordbok over det norske folkemålet og det nynorske skriftmålet, Oslo
- Raknes, Ola 1939-1942: *Fransk-norsk ordbok*, Oslo
- Rytter, Henrik 1932-1933: *Skodespel / William Shakespeare; i norsk umskrift ved Henrik Rytter*, Oslo
- Skard, Matias 1939: *Nynorsk ordbok for rettskriving og litteraturlesnad*, Oslo
- Torp, Alf 1915-1919: *Nynorsk etymologisk ordbok*, Kristiania
- Aasen, Ivar 1850: *Ordbog over det norske Folkesprog*, Christiania
- Aasen, Ivar 1873: *Norsk Ordbog : med dansk Forklaring*, Christiania

Notar

1. Band 5 blir utgjeve i 2005.