

En samlet ordbokpolitikk etter 2014

Rapport fra et seminar 20. mars 2013

Forord

Den 20. mars 2013 arrangerte Språkrådet et seminar i sine lokaler i Oslo om en samlet ordbokpolitikk i Norge etter 2014. Formålet med seminaret var å få best mulig grunnlag for et initiativ om ordbokpolitikk fra Språkrådets side.

Den språkpolitiske verdien av ordbøker og leksikon er omtalt slik i St.meld. nr. 35 (2007–2008) *Mål og mening*: ”Alle former for ordbøker og leksikalske oppslagsverk utgjør en viktig del av den totale skriftkulturelle infrastrukturen i eit språksamfunn. Alt dette er på sitt vis med på å styrkje skriftkulturen spesielt og nasjonalspråket generelt.” Året 2014 er et veiskille fordi det store prosjektet Norsk Ordbok 2014 blir avsluttet, samtidig som Det Norske Akademis Store Ordbok, NAOB, etter en tidligere plan også skulle være ferdig.

Spørsmål vi ville ta opp på seminaret, var blant andre: Hvilke behov ser vi i framtiden for ordbøker og lignende produkter? Hvordan skal ordbøkene være? Hva skal til for å dekke behovene? Hva skal det offentlige ha ansvaret for, og hva bør private ta seg av? Hvilke oppgaver skal Språkrådet ha?

Vi ønsket informasjon og synspunkter om mange typer ordbøker vi kjenner: de store, dokumenterende ordbokverkene, standardordbøkene *Bokmålsordboka* og *Nynorskordboka*, ordbøker for samisk og for språk som Språkrådet har fått et særskilt ansvar for – nasjonale minoritetsspråk og norsk tegnspråk, andre tospråklige ordbøker mellom norsk og andre språk, både de store skolespråkene og nyere minoritetsspråk – i bokform og digital form.

Representanter for ulike ordbokmiljøer ble invitert til å holde innlegg, og på slutten av dagen var ordet fritt for alle framfødte. Det ble et interessant og innholdsrikt seminar som berørte mange av spørsmålene ovenfor. Denne rapporten kan ikke yte rettferdighet overfor alle de små og store innleggene som kom i løpet av dagen. Vi har funnet det riktig å begrense den til å inneholde skriftlige versjoner av de inviterte innleggene. Disse er plassert i den rekkefølgen de ble holdt, slik det framgår av det gjengitte programmet for dagen. Ett av de inviterte innleggene (Austigard) forelå ikke i skriftlig form til produksjonen av denne rapporten.

I det frie ordskiftet ble det blant annet påpekt at mange av innleggene hadde en historisk og tilbakeskuende karakter, og redegjorde for ordbokprosjekter som var fullført eller nærmest fullført. Ved at rapporten begrenses til de inviterte innleggene, stiller vi oss igjen åpne for slik kritikk. Vi mener likevel at en viss kunnskap om situasjonen i dag og historien fram mot den er nyttig for den som ønsker å gjøre seg opp sin egen oppfatning.

På seminaret ble det ikke lagt skjul på at det finnes ulike oppfatninger om hvordan ordbokarbeidet best skal organiseres framover. Slik uenighet finner vi blant annet i spørsmålet om visse sider ved det planlagte prosjektet Bokmålets og riksmålets ordbokverk. Det er likevel vårt inntrykk etter seminaret at forståelsen av situasjonen er såpass felles for de fleste aktørene at vi i Språkrådet mener det er grunnlag for å gå vidare med et initiativ for en samlet ordbokpolitikk etter 2014.

Oslo, august 2013

Arnfinn Muruvik Vonen
direktør, Språkrådet

Innholdsfortegnelse

Forord	2
Innholdsfortegnelse	3
Arnfinn Muruvik Vonen: En samlet ordbokpolitikk?.....	5
Øystein Eek: En samlet ordbokpolitikk etter 2014?.....	7
Anna Helga Hannesdóttir: Ordbokspolitik i Sverige, Danmark, Island och Finland – en översikt	13
Ruth Vatvedt Fjeld: Leksikografisk bokmålskorpus (LBK 2013) og Bokmålets og riksmålets ordbokverk (BRO).....	18
Gisle Andersen: Korpusleksikografi i Norge: fra mønsterkorpus til monsterkorpus	21
Oddrun Grønvik: Kva slags språksamlingar treng Noreg etter 2014?	24
Åse Wetås: Ressursar og kompetanse i Norsk Ordbok 2014-prosjektet	28
Olaf Husby: Om tilstanden og behovene for ordbøker for innvandrere	31

Ordbokpolitikk i Norge etter 2014

Seminarprogram

20. mars 2013

Observatoriegata 1 B, Oslo

- 1000–1045 **En samlet ordbokpolitikk?**
Arnfinn Muruvik Vonen, Språkrådet
- 1045–1100 **Kommentarer til innledningsforedraget**
Øystein Eek, Kunnskapsforlaget
- 1100–1115 **Kommentarer til innleiingsforedraget**
Edmund Austigard, Det Norske Samlaget
- 1125–1210 **Ordbokspolitik – hur kan den se ut?**
Om situationen för lexikografisk verksamhet i Sverige och på Island
Anna Hannesdóttir, Göteborgs universitet
- 1255–1325 **Leksikografisk bokmålskorpus (LBK2013) og Bokmålets og riksmålets ordbokverk (BRO)**
Ruth V. Fjeld, Universitetet i Oslo
- 1325–1355 **Hvordan sikre jevn tilgang til pålitelige data om norsk språk etter 2014?**
Gisle Andersen, Norges handelshøgskole
- 1405–1435 **Korleis sikre jamn tilgang til pålitelege data om norsk språk etter 2014?**
Oddrun Grønvik, Norsk Ordbok 2014
- 1435–1500 **Norsk Ordbok: Ressursar i 2014-prosjektet og fulldigital integrering av alfabetstrekket a–h**
Åse Wetås, Norsk Ordbok 2014
- 1500–1525 **Om tilstanden og behovene for ordbøker for innvandrere**
Olaf Husby, Norges teknisk-naturvitenskapelige universitet (NTNU)
- 1525–1600 **Åpen mikrofon**

En samlet ordbokpolitikk?

Arnfinn Muruvik Vonen
Språkrådet

På vegne av Språkrådet vil jeg få ønske alle velkommen til dette seminaret, som har som hensikt å belyse ordboksituasjonen i Norge nå som vi nærmer oss 2014. 2014, grunnlovsjubileumsåret, er året da det største igangværende ordbokprosjektet i landet, Norsk Ordbok 2014, planlegges slutført, og året da et annet stort prosjekt, Det Norske Akademi's Store Ordbok, var planlagt slutført ifølge en tidligere plan. Vi ønsker å få fram en debatt om hvilke ordbokbehov som fortsatt gjenstår i landet når det store krafttaket som Norsk Ordbok 2014 er, blir historie. Ved å invitere innledere fra ulike ordbokfaglige miljøer ønsker vi å se spørsmålet fra mange ulike perspektiver.

Ordbøker er i alle skriftkulturer en nødvendig forutsetning for språklig demokratisk deltakelse. I gode ordbøker kan alle finne igjen sitt eget språk, og med ordbøker kan alle få hjelp til å bli enda bedre og tryggere språkbrukere. De fleste språkbrukerne greier seg gjerne med ordlister eller mindre ordbøker, men disse blir bare gode hvis de er utviklet fra store språksamlinger og kan bygge på moderne, oppdaterte definisjonsordbøker.

Språkrådet er statens fagorgan i språkspørsmål, og våre vedtekter slår fast både at vi skal observere og analysere aktuelle utviklingstrekk i hele det norske språksamfunnet, og at vi på eget initiativ bør legge til rette for tjenlig samarbeid med både private og offentlige aktører innenfor ulike samfunnssektorer. Videre har vi ansvar for å ta språkpolitiske initiativer overfor relevante myndigheter.

De språkpolitiske målene for Norge er gitt i relevante styringsdokumenter, først og fremst språkmeldingen, St.meld. nr. 35 (2007–2008) *Mål og mening. Ein heilskapleg norsk språkpolitikk*. I meldingen heter det (kap. 8.3.1): ”Alle språk som skal haldast i hevd og utvikla seg som levande kulturspråk, må kunna støtta seg på ein solid dokumentasjon av den språklege tradisjonen og av korleis språket blir brukt i samtida.” Også i innstillingen om språkmeldingen frå Stortingets familie- og kulturkomité (kap. 2.4.1) understrekes behovet for språklig dokumentasjon: ”Språkendingar – i tale og skrift – skjer rett nok ofte som generasjonssprang, men det er svært viktig at ein sikrar språkleg dokumentasjon og setter inn tiltak som på ulike vis kan styrkje språket.”

For at Språkrådet skal kunne utføre sitt oppdrag med å følge med på faktisk språkbruk, må vi ha tilgang til god språklig dokumentasjon.

Jeg vil forsøksvis gjøre en inndeling av ordbokfeltet i Norge i dag, i det håp at det kan hjelpe til en ryddig og oversiktlig debatt:

1. Ettspråklige allmennordbøker
 - a. Skoleordlister og -ordbøker
 - b. Standardordbøker og ettbinds definisjonsordbøker
 - c. Store vitenskapelige ordbøker
2. Tospråklige ordbøker
3. Fagordbøker

Språkrådet har et særlig ansvar for skoleordlistene og skoleordbøkene, da vi etter opplæringsloven skal godkjenne disse ordlistene og ordbøkene. Vi har også et særlig ansvar for at rettskrivnings- og definisjonsordbøkene *Bokmålsordboka* og *Nynorskordboka* formidler den til enhver tid gjeldende rettskrivningsnormen. Når det gjelder de øvrige kategoriene av ordbøker, er vår rolle i utgangspunktet mer begrenset. Med vår kompetanse på feltet kan vi likevel følge med på feltet og i kontakt med andre aktører vurdere hva som er behovene for videre utvikling.

Prosjektene Norsk Ordbok 2014 og Det Norske Akademis Store Ordbok er spesielt interessante å følge med på fordi de er så omfattende og betydningsfulle for dokumentasjonen av norsk språk. Det samme gjelder det planlagte prosjektet Bokmålets og riksmålets ordbokverk. Vi har derfor valgt å sette av en ganske stor del av tiden på seminaret til å se nærmere på disse prosjektene. Ikke desto mindre ønsker vi også å trekke opp et bredere felt av behov, der også tospråklige ordbøker mellom norsk (bokmål og nynorsk) og en lang rekke andre språk er meget viktige. I lys av det stadig mer synlige språklige mangfoldet i Norge melder det seg også behov som tidligere ikke ble like sterkt vektlagt, som at tospråklige ordbøker må kunne brukes av primærbrukere av begge aktuelle språk.

Det er et gjennomgående trekk ved de språkene som ved siden av norsk har en lang historie her i landet, at de mangler tilfredsstillende leksikografisk dokumentasjon: de samiske språkene, de nasjonale minoritetsspråkene (kvensk, romani og romanes) og norsk tegnspråk. Ved å styrke de fagmiljøene som allerede er i gang med faglig kvalitetssikrede ordbokprosjekter, kan man bidra på en synlig og viktig måte til styrking av språk som vi har et spesielt ansvar for i Norge.

Når det gjelder fagordbøker, utgis det hvert år et lite antall slike på kommersielle forlag i Norge. De fleste av dem er tospråklige, for det meste mellom bokmål og engelsk. Dette skaper et økende behov for terminologi også på nynorsk. I Språkrådet arbeider vi med å legge til rette for terminologiarbeid i Norge generelt, og også naturligvis på nynorsk.

Det er vanskelig å tenke seg kvalitetsordbøker uten språklig dokumentasjon som ligger til grunn, som korpus og andre språklige samlinger. Av den grunn må også spørsmålet om språklig innsamlings- og systematiseringsarbeid stå på dagsorden på et seminar som dette.

Med dette vil jeg invitere publikum – og rapportlesere – inn i debatten.

En samlet ordbokpolitikk etter 2014?

Kommentarer til Arnfinn Muruvik Vonens foredrag på Språkrådets ordbokseminar 20.03.2013

Øystein Eek
Kunnskapsforlaget

1 Innledning

1.1 Vårt hovedbudskap

Vi ønsker oss en samlende ordbokpolitikk!

Målet må være å tilby en best mulig dokumentasjon av både norsk og fremmedspråk, av de språkene som i tillegg til norsk har offisiell status i Norge, og av minoritetsspråkene, og forbindelsen mellom norsk og disse språkene.

Anglifiseringen, ikke minst i arbeidslivet, går raskere enn vi liker å tenke på. Dette har mange bivirkninger, og gjør det svært viktig og nødvendig å dokumentere norsk som et komplett språk, for alle aktuelle samfunns- og fagområder.

Vi oppfatter formuleringen “en samlet ordbokpolitikk”, og siktemålet med Språkrådets initiativ slik:

- Å beskrive hele feltet
- Å klargjøre Språkrådets rolle i ordboksektoren
- Å peke på ønskelige tiltak

Men: En samlet ordbokpolitikk vil ha bedre sjanse for å lykkes hvis den også er en samlende ordbokpolitikk.

En samlende ordbokpolitikk skal

- ivareta behovene til alle brukergrupper, uansett målform og språk
- ivareta interessene til dem som driver seriøst med ordbøker
- ta vare på kompetanse, engasjement og satsing

Hovedspørsmålet i innledningsforedraget er “Hva kan og bør gjøres for å sikre at det fortsatt blir laget gode og nyttige ordbøker i Norge etter 2014?” Vi er glade for at problemstillingen reises så vidt som dette.

1.2 Kunnskapsforlagets hovedmål

- Vi vil tilby et bredt utvalg av digitale ordbøker på alle plattformer som er drivverdige.
- Vi vil tilby trykte ordbøker som brukerne etterspør.
- Vi vil ha en egen redaksjon.
- Vi vil ha god kommunikasjon med kundene, også om faglige spørsmål.
- Vi vil drive seriøs rettighetsforvaltning.
- Vi vil ha gode relasjoner til Språkrådet, de akademiske miljøene og Det Norske Akademi, og selvsagt også til andre samarbeidspartnere og aktører som NTB, NPK og oversetterforeningene.
- Vi vil fungere som en selvstendig, selvberende økonomisk enhet.

(Kunnskapsforlaget er ett av tre firmaer i Norge som lever av å utgi og selge ordbøker. Egentlig er det bare to firmaer som lever av å produsere ordbokinnhold og publisere og selge dette, nemlig Kunnskapsforlaget og Clue. iFinger har utviklet en teknologisk løsning og et markeds- og salgsapparat som selger denne løsningen med innhold lisensiert fra diverse ordbokforlag. For andre forlag som utgir ordbøker, er dette bare en mindre del av virksomheten.)

2 Kommentarer

Innledningsforedraget og mine kommentarer omhandler to hovedtemaer: ordboksituasjonen generelt og de store dokumentasjons- og publiseringsprosjektene som får offentlige tilskudd. Andre aktuelle temaer, bl.a. ordbøker for de andre offisielle språkene i Norge og for minoritetsspråkene, kan jeg dessverre ikke komme inn på. Jeg må også la spørsmålet om de store samlingene på Universitetet i Oslo ligge.

Kommentarene til innledningsforedraget går inn på følgende temaer:

- Kulturdepartementet og dokumentasjonsordbøkene
- Språkrådets funksjoner
 - i ordbokpolitikken
 - overfor NAOB
 - overfor Akademiet
- BRO og NAOB
- Universitetenes oppgaver
- Norsk Ordbok etter 2014
- Ingenting er gratis
- Forlagenes rolle
- Offentlig sektor som kunde
- Finansiering av ordbokprosjekter

2.1 Kulturdepartementet og dokumentasjonsordbøkene

Staten må ha som ordbokpolitikk at begge målformene skal ha et nasjonalt ordbokverk: NO 2014 for nynorsk og (utvalgte) dialekter, BRO/NAOB for bokmål/riksmål.

Det må legges til rette for digital utgivelse av NAOB og NO2014.

Tilstrekkelige midler må følge denne politikken, finansiering av den ene målformen må ikke gå på bekostning av den andre.

Tilskuddene må også omfatte driftsmidler for nettstedene, ettersom tjenestene skal være fritt tilgjengelige.

2.2 Språkrådets funksjoner i ordbokpolitikken

Språkrådet bør

- holde kontakt med aktørene og være lydhørt for synspunkter og innspill
- vurdere behov og komme med ønsker
- anbefale prosjekter overfor bevilgende myndigheter
- forvalte normen for bokmål og nynorsk

- også interessere seg for og informere om prosjekter og utgivelser som følger egne operative normer
- oppdatere retningslinjer og rutiner for å godkjenne rettskrivningsordlister for norskundervisningen, og så iverksette dem
- arbeide for norsk språk som helhet – bokmål og nynorsk side om side. Målloven skaper et dilemma her, men det må være mulig å løse med et godt kommunikasjonsarbeid

Språkrådet bør ikke

- drive eget ordbokarbeid og ha egne rettigheter. Språkrådet mangler apparat og stimuli til å stille med dette. Den historiske alliansen med Universitetet i Oslo / ILN har kanskje ikke vært bare heldig

2.3 Språkrådet og NAOB

Styrets engasjement på vegne av NAOB, BRO og bokmålet i 2011/12 har vært viktig. Det er også viktig at Språkrådet engasjerer seg i fortsettelsen.

Ut fra Arnfinn Vonens helhetstilnærming i innledningsforedraget bør nærheten til Norges største ordbokforlag betraktes som en styrke for NAOB, både når det gjelder prosess og produkt. Også dette bør kunne sies eksplisitt.

Vi merker oss formuleringen «et godt faglig fundert BRO», som prosjektet absolutt skal leve opp til.

2.4 Språkrådet og Akademiet

Blant flere andre jubileer i Språkrådet 2013 er det grunn til å nevne at også Det Norske Akademi for Sprog og Litteratur har 60-årsjubileum, som skal feires til høsten.

Akademiet har en viktig funksjon i norsk språkrøkt og kan være en god støttespiller for Språkrådet i det langsiktige normeringsarbeidet for bokmål.

En viktig premis for dette arbeidet er at Språkrådet og Akademiet har god kontakt og gjensidig respekt.

2.5 BRO og NAOB

Kulturdepartementets visjon om samarbeid innen bokmål/riksmål mellom kompletterende miljøer har blitt realisert på beste vis gjennom BRO.

Prosjektredaksjonen bør kunne utvides med redaktører fra Norsk Ordbok 2014-miljøet hvis midler stilles til disposisjon. Dagens NAOB-redaksjon har høy kompetanse, men er for liten. (De seks veteranene i redaksjonen har til sammen mer enn 220 års erfaring med allmenn- og fagleksikografisk arbeid. To av redaktørene har doktorgrad, prosentvis er vel dette en like høy andel som i Norsk Ordbok 2014.)

BRO og NAOB må styres av dem som har de historiske og språkideologiske hovedinteressene i prosjektet. Slik kommer prosjektet på best mulig måte språkstyrkingen i landet til gode, ref. innledningsforedraget.

NAOBs redaksjonelle metodikk og IT-system har vært presentert på flere seminarer og konferanser. Det er betydelig interesse i miljøet i Oslo for disse spørsmål, og NAOB vil gjerne arrangere et nytt seminar for interesserte om dette.

2.6 Universitetenes oppgaver

Universitetene bør først og fremst drive forskning og undervisning, og formidling av forskningsresultater.

Det er ikke nærliggende at universiteter forvalter og drifter publiseringsprosjekter for allmennheten.

Det er ønskelig at flere miljøer blir involvert i de store dokumentasjonsprosjektene, gjerne med delprosjekter.

Rettighetsforvaltningen ved Universitetet i Oslo har vist seg å være lite tilfredsstillende. Dette skaper langvarige og uoversiktlige prosesser som er lite produktive.

Den redaksjonelle driften og publiseringen av Bokmålsordboka og Nynorskordboka bør diskuteres. Sammenlignet med normale forlagsaktiviteter er disse blitt lite redaksjonelt utviklet siden lanseringen, trolig på grunn av få eksterne stimuli.

2.7 Norsk Ordbok etter 2014

Vi mener, som de fleste som har ytret seg om dette, at de viktigste oppgavene må være

- digitalisering av hele alfabetet
- oppdatering
- drift av digital utgave
- permanent fornyelse

Det må finnes en redaksjon på permanent basis, og nødvendig IT-støtte.

Men: Dette må ikke gå på bekostning av finansieringen av bokmålets dokumentasjon.

For brukerne er Norsk Ordbok først og fremst en ordbok, med redaksjonelt tilrettelagt informasjon. Alt som ligger bak, av kildemateriale og forarbeider, er vel mest av interesse for forskere og andre spesialister. (Åse Wetås fremholdt i et senere foredrag at det er stor interesse blant brukerne av nettstedet også av de bakenforliggende ressursene, men dette forteller kanskje at de foreløpige brukerne er mer enn vanlig språk- og tradisjonsinteresserte.)

Vi mener det ville være fornuftig å la utgivelsen bli lagt til en instans som også driver med andre (ordbok)utgivelser.

Det er vel også aktuelt å vurdere om abonnentene burde tilbys oppdaterte versjoner av de første bindene, jf. Norsk Riksmålsordboks tilleggsbind fra 1995.

Redaksjonelt kunne Norsk Ordbok gjerne forankres i et forlagsmiljø som også ellers driver med ordbøker.

2.8 Ingenting er gratis

Ordbokpublisering og -salg forvaltes i prinsippet av privat sektor i Norge. Dette har fungert bra i over 100 år, og kan fortsatt fungere bra, gitt tilfredsstillende rammevilkår.

All produksjon og publisering er forbundet med kostnader. Det offentlige bør forholde seg til gjeldende forretningsmodeller.

Digitale produkter kan i visse tilfeller gjøres gratis for sluttbrukerne, men dette forutsetter synlige og kontrollerbare driftstilskudd på vegne av samfunnet, eller andre former for storkundeavtaler.

Det vanlige annonsemarkedet fungerer ikke for denne type utgivelser.

2.9 Forlagenes rolle

Kunnskapsforlaget, Samlaget, Cappelen/Damm, Vega, Fagbokforlaget o.a. konkurrerer om det lille markedet for trykte ordbøker, som er blitt kraftig redusert over hele den vestlige verden de siste 5–10 år. Det er neppe noen som alvorlig overveier å legge dette under staten. Rettighetsforvaltning er et kjerneområde i forlagene, og er viktigere enn noensinne. Seriose aktører ivaretar opphavsmennenes interesser på en profesjonell måte.

Clue, Kunnskapsforlaget og iFinger konkurrerer om det lille digitale markedet. Staten eller andre offentlige virksomheter bør avstå fra å forsøke å legge denne virksomheten under seg, jf. NDLAs (Norsk digital læremiddelarenas) varslede ordboksatsing.

Forlagene (i vid forstand) må kunne forvalte sine porteføljer i samsvar med kundenes behov også på lengre sikt, og med inntekter som gir grunnlag for fortsatt utvikling av IT, kundefølgning og redaksjon.

2.10 Bokmålsordboka og Nynorskordboka

Det er uheldig at disse ordbøkene ikke har vært underlagt samme typer forlagsavtaler som andre titler. Som en konsekvens av det som er sagt allerede, ønsker vi at rettighetssituasjonen normaliseres slik at Samlaget og Kunnskapsforlaget har vanlige vilkår for utgivelse. Også de redaksjonelle prosesser må normaliseres.

2.11 Offentlig sektor som kunde

Gjennom skoleverk og offentlige institusjoner er en stor andel av ordbokmarkedet i dagens Norge knyttet til offentlige innkjøp.

Det er da viktig at disse innkjøp skjer i et reelt marked, med reell konkurranse, med vurderingskriterier som har en balanse mellom kvalitet, funksjonalitet og pris, og til priser som gir bærekraft for leverandørene.

Anbudsinstituttet er uheldig for åndsverk. Prisene på det offentlige, digitale markedet, særlig skoleverket, presses altfor langt ned, slik at skolemarkedet må subsidieres av annet salg og med den følge at den redaksjonelle utviklingen hemmes. Dette er en usunn situasjon. Det offentlige må innse at leverandørene må ta seg noe bedre betalt hvis kvaliteten på produktene skal opprettholdes over tid.

2.12 Finansiering av ordbokprosjekter

Ordbokfloraen, særlig den tospråklige, forvaltes av idealister som jevnt over er kommet godt opp i årene.

Fremtidige innholdsleverandører vil ikke nøye seg med royalty, men vil ha betalt for arbeidet. Det må være mulig å få (del)finansiering av prosjekter som er etterspurt av Språkrådet og andre, men som er for risikable for utgiverne.

Dette er den billigste og mest effektive måten å opprettholde ordbøker til og fra fremmedspråk på.

Ingen andre enn forlagene vil og kan ta ansvar for at det skal finnes slike utgivelser på et faglig høyt nivå.

Søknader om støtte bør vurderes av en mest mulig uavhengig kommisjon.

De vanlige forretningsmodellene bør gjelde.

3 Oppsummering: En samlende ordbokpolitikk

Norge trenger en samlende ordbokpolitikk, som må

- bygge videre på eksisterende arbeidsdeling og forretningsmodeller
- legge til rette for bredt leksikografisk arbeid også med bokmål
- gi mulighet for økonomisk støtte til utvalgte prosjekter
- arbeide for økonomisk forståelse og forståelse av rettighetenes betydning

Ordbokspolitik i Sverige, Danmark, Island och Finland – en översikt

Anna Helga Hannesdóttir
Göteborgs universitet

Sammanfattning

Efter den genomgång av ordbokspolitik i Sverige, Danmark, Island och Finland som här redovisas, kan följande framträdande komponenter urskiljas: Det nationella ordboksprojektet, rättstavningsordböcker, dialektordböcker, terminologiarbete, minoritetsspråksstöd och ordböcker som ett led i integrationsarbetet.

I ett historiskt perspektiv har stora, enspråkiga ordböcker betraktats som nationellt viktiga projekt. De ambitiösa ordboksarbeten som startade runt sekelskiftet 1900 är fortfarande, på olika sätt, centrala för respektive lands lexikografiska verksamhet, antingen genom att verksamheten fortfarande pågår eller genom att de miljöer som då etablerades fortfarande består och utgör nationella centra för lexikografisk kompetens och produktion.

Idag ansvarar vanligen de statligt finansierade språkvårdsorganen för de rättstavningsordböcker där den officiella skriftnormen förmedlas. Genom språkteknologiska satsningar görs befintliga ordböcker tillgängliga på nätet och sedelsamlingar och annat lexikografiskt material digitaliseras för att bli åtkomligt. Uppbyggnad, utveckling och underhåll av korpusar och sökverktyg ingår i uppdraget till de offentliga aktörerna liksom stöd till terminologiarbete och dialektordböcker. Ett område som det satsas på i Sverige och Finland är språkplaneringsarbetet för de nationella minoriteternas språk, där även lexikografiskt arbete i olika utsträckning ingår. Också utarbetande av ordböcker över aktuella invandarspråk kan ingå, som ett led i en integrationspolitik.

Inledning

En kortfattad översikt över ordböckers plats i de nordiska ländernas offentliga språkpolitik kan naturligtvis inte bli annat än ytlig. Här är det främst den svenska ordbokspolitiken som står i fokus, medan situationen på Island och i Danmark och Finland skisseras mera kortfattat. Förlagens roll utelämnas helt. Istället är det synen på ordboken som en nationell resurs och ansvaret för denna resurs som redovisas.

Sverige

Den enspråkiga svenska lexikografiska verksamheten präglas av det uppdrag Gustav III gav Svenska Akademien (SA) vid dess inrättande år 1786. Akademien skulle bl.a. utarbeta en svensk ordbok. Först framåt mitten av 1800-talet kom arbetet med ordboken igång på allvar. När SA år 1869 beslutar att ge ut en ordlista, ett slags ram till ordboken, tas ett viktigt steg i konsolideringen av den roll Akademien spelar i dagens svenska ordbokspolitik. Första upplagan av *Svenska Akademiens ordlista*, SAOL, publicerades 1874 och följdes i rask takt av nya upplagor. När 6 upplagan kom ut 1889 gjordes den till officiell norm för skolornas rättstavning och vid stavningsreformen 1906 (den senaste som svenskan har genomgått) bekräftades denna status i ett kungligt cirkulär. Akademiens konservativa språksyn avvek

emellertid i vissa avseenden från det bruk som med tiden etablerades men 1951 stadfästes återigen ordlistans normerande status.¹ I sekreterarens förord till 13 upplagan presenteras SAOL som

[...] en samling rekommendationer, som i stor utsträckning grundas på det etablerade bruket och som därutöver bestäms av en strävan att anpassa nya tillskott i språket till rådande svenska normer för stavning och böjning, ordbildning och ordval. (Engdahl 2006:V)

Arbetet med 14 upplagan pågår nu vid Lexikaliska institutet, Institutionen för svenska språket vid Göteborgs universitet.

Första häftet av den ordbok som SA fick i uppdrag att åstadkomma publicerades 1893. *Ordbok över svenska språket utgiven av Svenska Akademien* (SAOB) är en produkt av det sena 1800-talets syn på språkvetenskap: den omfattar det svenska ordförrådet från 1521. Förutom ordens betydelse redovisas deras etymologi och utveckling. Band 36, *tynga–utsudda*, avslutades 2012. Sista bandet planeras utkomma 2017. SA har med 2000-talets inträde befest sin ställning som den enspråkiga svenska lexikografins huvudman. 2009 utkom *Svensk ordbok utgiven av Svenska Akademien* (SO), en ordbok över dagens svenska. Såväl SAOL som SAOB finns i elektroniska utgåvor, SAOL som faksimil på Akademiens hemsida samt som app och SAOB som sökbar databas. De elektroniska utgåvorna är gratis tillgängliga.

SA är en fristående kulturinstitution som, med utgångspunkt i det nationella ordboksprojekt som grundaren föreskrev, nu uppbär huvudmannskapet för den lexikaliska beskrivningen av svenskan. Genom att med egna medel finansiera sina uppdrag till ordboksredaktionerna i Lund (SAOB) och vid Göteborgs universitet (SAOL och SO) och med en plan för det fortsatta arbetet på SAOB efter 2017 samt för återkommande upplagor av SAOL och SO ter sig framtiden för de enspråkiga ordböckerna över den moderna standardsvenskan ganska stabil.

Lexikografisk verksamhet i statens regi inriktas huvudsakligen på ordböcker mellan svenska och aktuella invandarspråk, de s.k. lexinordböckerna. En gemensam svensk källspråksdatabas kompletteras med ekvivalenter och annat material på det aktuella målspråket. Syftet med ordböckerna är att de skall användas i undervisningen i svenska som andraspråk. Av det tjugotal ordböcker som publicerats finns de flesta i digital form, gratis tillgängliga på svenska Språkrådets hemsida. För närvarande pågår arbetet med ordböcker över tre romska språk samt thai. Förutom lexinordböckerna bidrar staten till terminologiverksamheten vid Terminologicentrum samt till arbetet med dialektordböcker vid Institutet för språk och folkminnen. Dessutom har vissa tvåspråkiga ordböcker bekostats med statliga medel, exempelvis de tryckta ordböckerna mellan nederländska och svenska och den svenska målspråksdelen av de digitala ISLEX-ordböckerna.

Trots att alltså en hel del svenskt ordboksarbete bekostas med offentliga medel är det svårt att urskilja en samlad svensk ordbokspolitik. Verksamheten vid den statliga myndigheten Institutet för språk och folkminnen och därtill hörande Språkrådet sorterar under Kulturdepartementet, det terminologiska arbetet uppbär stöd från Näringsdepartementet och arbetet med svenskan i ISLEX finansierades helt av Utrikesdepartementet.

¹ Uppgifterna om arbetet med SAOL har i stor utsträckning hämtats i de olika uppsatserna i *SAOL och tidens flykt* <spraakdata.gu.se/saolhist/>.

Framtiden för svensk lexikografi diskuterades hösten 2012 vid ett symposium som anordnades av Institutionen för svenska språket vid Göteborgs universitet. Följande problem lyftes fram:

(1) Utbildningen av framtidens lexikografer. I Sverige har många av ordboksförlagen lagt ned sina ordboksredaktioner. Arbetsmarknaden för lexikografer har minskat drastiskt vilket får följderna för universitetens språkämnen. Också förlagens roll som ordbokspolitiska aktörer inom främst tvåspråkig lexikografi har ändrats radikalt. Den finansiering av ordboksprojekt som förlagen bidragit med behöver kompenseras när intäkterna av ordboksförsäljningen rasar.

(2) Lexikografins forskningsförankring behöver stärkas och lexikografen måste anpassas till de nya publiceringsformerna. De möjligheter som nya publikationsformer öppnar kräver fördjupad lexikografisk kompetens och nya arbetsmetoder. Samtidigt behöver lexikografins status som akademisk verksamhet höjas.

(3) Lexikografi bör betraktas som nationell infrastruktur som behöver stöd för underhåll och utbyggnad.

Danmark

Även i Danmark genomfördes ett stort ordboksarbete av formatet nationellt projekt: *Ordbog over det danske Sprog* (1918–1956) med supplement (1992–2005). Ordboken, som omfattar det danska ordförrådet från 1700 till 1950, uppdrogs åt Det Danske Sprog- og Litteraturselskab (DSL). Även en stor ordbok över danskan från 1955, *Den Danske Ordbog*, har utarbetats vid DSL. Ordboken kom ut i tryck 2003–2005. Båda dessa arbeten har digitaliserats och finns nu gratis tillgängliga på nätet. DSL är, liksom Svenska Akademien, en fristående kulturinstitution men till skillnad från den sistnämnda finansieras DSL:s verksamhet delvis med offentliga medel och externa fondmedel.

En annan aktör inom den offentliga lexikografiska verksamheten är Dansk Sprognævn. Nämnden redigerar och ger ut den rättskrivningslista som utgör den officiella normen för rättskrivningen. Den har mandat att genomföra smärre ändringar i rättskrivningen, men ändringar av principiell karaktär skall godkännas av ansvariga ministrar. I uppdraget ingår också att samarbeta med terminologiorgan och med ordboksredaktörer.

Island

Första gången isländska behandlades lexikografiskt var på 1600-talet, i isländsk-latinska arbeten utgivna i Danmark och i Sverige. Trots denna tidiga start är det inte förrän i samband med arbetet för självständighet som man kan tala om en isländsk ordbokspolitik. Det första initiativet till en enspråkig historisk ordbok över isländskan från år 1540 togs 1918, samma år som Island fick självstyre. Satsningen fick statlig finansiering från 1947. Någon sådan ordbok har det emellertid inte blivit, men det insamlade materialet har digitaliserats och gjorts tillgängligt på nätet. 1957, tretton år efter att landet blev självständigt, inleddes arbetet med en ordbok över den moderna isländskan. Ordboken publicerades 1963, nya upplagor 1983, 2002 och 2007. Arbetet finansierades från början med offentliga medel, men med tiden har beroendet av fonder och förlagsstöd blivit allt större.

Den isländska rättskrivningen regleras i ett offentligt direktiv och presenteras i en rättskrivningsordbok som utarbetas av isländska språknämnden i samarbete med ordboksredaktionen vid Islands universitet. Ändringar av principiell karaktär skall godkännas

av ansvarig minister. I språknämndens uppdrag ingår även att samarbeta med och vara sekretariat för enskilda termgruppers arbete.

Finland

Finska statens främsta språkpolitiska aktör är Institutet för de inhemska språken. Institutets arbete omfattar finsk och svensk språkvård, namnvård, finska och finlandssvenska dialekter m.m., och institutet skall även samordna språkvården av romska, samiska och teckenspråk. Också ordboksarbete samt forskning med anknytning till ordboksarbete ingår i uppdraget. Vid institutets ordboksavdelning bedrivs såväl vetenskapligt grundad lexikografi som lexikografisk forskning och där produceras enspråkiga ordböcker över finska, finlandssvenska och dialekter samt tvåspråkiga ordböcker. Ordboksavdelningen förfogar över egna databaser och textkorpusar som kontinuerligt uppdateras och utvecklas.

För svenskan i Finland är det i princip de sverigesvenska normerna, såsom de presenteras i SAOL, som gäller för rättskrivning och ordböjning. Finlandssvensk språkvård omfattar därför främst finlandssvenskan i relation till sverigesvenskan och de avvikelser från denna som utvecklats i det finlandssvenska riksspråket. Institutet ger även ut ordboken *Kielitoimiston sanakirja*, en storordbok som kan sägas förmedla de officiella riktlinjerna för finskans användning.

Ordböcker i skolans styrdokument

Vilken betydelse samhället tillmäter ordböcker kan utläsas av huruvida de behandlas i styrdokumentet för skolundervisningen. En undersökning av styrdokumentet för den danska, norska och svenska skolan visar att situationen varierar stort (Rødningen & Karlsen 2008). I fokus står den norska skolan som inte behandlas vidare här. Överlägset störst ambitioner när det gäller att lära eleverna att använda ordböcker uttrycks i de danska styrdokumentet, sammanfattat som följer:

Danskane er definitivt best i klassen. Gjennom heile læreplandokumentet for morsmålsfaget i grunnopplæringa er det lagt vekt på at elevane skal bruke ordbøker aktivt. I undervisningsrettleiinga for læraren er det eksplisitt streka under at elevane skal ha tilgang til ulike typar ordbøker, og at læraren sjølv skal vere eit føredøme for elevane ved sjølv å bruke ordbok aktivt. (Rødningen & Karlsen 2008)

Den svenska grundskolan fick nya kursplaner 2011. Medan de tidigare kursplanerna för svenska och svenska som andraspråk inte nämnde ordböcker alls skall nu elever i båda ämnena efter skolår 6 veta hur man använder ”ordböcker och andra hjälpmedel för stavning och ordförståelse” och vid grundskolans slut kunna använda dem. I de övriga språkämnen har emellertid det krav som tidigare ställdes inom engelskämnet – att eleverna efter skolår 9 skall ”kunna välja och använda hjälpmedel vid textläsning, skrivning och andra språkliga aktiviteter” – ersatts av formuleringen att eleverna skall ”ges förutsättningar att kunna använda olika hjälpmedel för lärande, förståelse, skapande och kommunikation”. Samma formulering återkommer i kursplanen för ämnet moderna språk <skolverket.se/kursplaner-och-betyg>.

Den isländska skolans styrdokument ingick inte i Rødningens och Karlsens undersökning. Där ges ordböcker och lexikala hjälpmedel tydligare status än i de svenska styrdokumentet. I årskurs 7 och 9 i isländska, isländska som andra språk samt moderna språk ska eleven kunna använda hjälpmedel såsom ordböcker, korpusar, rättstavningsprogram och sökmotorer.

Förutom i de danska kursplanerna förefaller ordböcker spela en marginell roll i språk-undervisningen såsom den bedrivs i Nordens skolor.

Ordböcker som infrastruktur

Uppbyggnad och underhåll av nationell infrastruktur ingår i myndigheternas åtaganden. De nordiska myndigheterna tycks ännu inte betrakta ordböcker som infrastruktur nödvändig för kunskapsutveckling och kommunikation. Trots nya möjligheter med nya verktyg och nya publikationsformer är lexikografiskt arbete fortfarande tidsödande och kompetenskrävande. De lexikografiska material som redan finns tillgängliga i olika former utgör en nationell tillgång. Utbyggnad, underhåll och utveckling av detta material borde vara en uppgift av samma dignitet som utbyggnad, underhåll och utveckling av annan nationell infrastruktur.

Av genomgången av de olika ländernas ordbokspolitik framgår de olika aktörernas betydelse. I Sverige och Danmark är kulturinstitutionerna oerhört viktiga för den lexikografiska beskrivningen av svenskan respektive danskan. I Danmark är emellertid arbetet vid DSL beroende av statliga anslag och av anslag från privata fonder. I Sverige spelar den från staten helt fristående Svenska Akademien en avgörande roll för den enspråkiga lexikografin. Genom sitt sätt att sköta det kungliga uppdraget fick 1786 framstå Akademien som en garant för en stabilitet som vida överträffar villkoren för den offentligt finansierade lexikografiska verksamheten i de övriga länderna.

Litteratur

Engdahl, Horace 2006: Förord. I: *Svenska Akademiens ordlista över svenska språket*. Stockholm.

Rødningen, Dagfinn & Knut E. Karlsen 2008: Leksikografi i pedagogisk perspektiv. Ein presentasjon av Norsk Ordbok for skoleelevar og studentar. I: *Nordiske studier i leksikografi* 9. Reykjavík. S. 371–381.

SAOL och tidens flykt 2009. Stockholm: Norstedts samt: <http://spraakdata.gu.se/saolhist>

Leksikografisk bokmålskorpus (LBK 2013) og Bokmålets og riksmålets ordbokverk (BRO)

Ruth Vatvedt Fjeld

Universitetet i Oslo

Leksikografisk bokmålskorpus (LBK) er et representativt korpus laget for leksikografiske formål som inneholder tekster fra perioden 1985 til i dag, og dermed dokumenterer ordforrådet i moderne bokmål. Et delmål er å bygge et korpus med minst 100 millioner ord fra ulike typer tekster og sjangre, balansert og vektet for leksikografiske formål.

I perioden 2000–2008 ble det samlet inn ca. 40 mill. ord, som var delmål 1. Dette korpuset ble ordklassetaggert med Oslo-Bergen-taggeren og lagt inn i Open Corpus Workbench, utviklet ved IMS i Stuttgart. Brukergrensesnittet Glossa, som korpuset presenteres i, er utviklet ved Tekstlaboratoriet ved Universitetet i Oslo.

Så vidt vi vet, var LBK2008 Norges første balanserte tekstkorpus. Av hensyn til tekstgiverne kunne ikke korpuset gjøres fritt tilgjengelig på nettet, da det bl.a. inneholder mange moderne og fortsatt salgbare tekster, men det er tilgjengelig for språkforskere som godkjenner noen få bruksregler.

Balanseringen var satt ut fra forskjellige hensyn: For det første ønsket vi en viss fordeling mellom forskjellige tekstsjangre, f.eks. romaner, lærebøker, offentlig informasjon, blogger. Dessuten ville vi dekke flest mulige fagfelt, som jus, medisin, realfag, humaniora, sport og fritid. Målet var å dekke omtrent alle typer tekster gjennomsnittlige språkbrukere møter i et relativt vanlig liv i det norske samfunnet, og vi støttet oss her på SSBs utredninger av folks lesevaner.

Vi fant imidlertid ut at 40 mill. ord var i minste laget for å gjøre de undersøkelsene av ordforrådet som vi ville, og søkte om midler til en utvidelse til 100 mill. løpeord, som er blitt vanlig standard for slike korpusressurser. Universitetet har fullfinansiert dette korpuset, som vi nå kaller LBK2013, og som har nådd dette målet i disse dager.

Målet for balansen var i 2008: 20 % periodika, 45 % sakprosa, 25 % skjønnlitteratur, 5 % TV-tekster og 5 % upublisert tekst. Ut fra endring i lesevaner og også med et sideblikk på hva slags tekster det var mulig å få tak i, har vi justert balansen noe i 2013, slik at periodika nå utgjør 10 %, sakprosa 45 %, skjønnlitteratur 35 %, TV-tekster 5 % og kategorien upublisert ble omdøpt til unormert, men utgjør fortsatt 5 %. Å skille mellom publiserte og upubliserte tekster er ikke lenger så greit, da mye tekst nå publiseres uredigert på Internett. Hovedskillet for vårt formål er å skille mellom tekster som er lest og redigert eller rettet av en normerende instans eller ikke, og vi har derfor valgt å kategorisere disse tekstene som unormerte, siden de er publisert og tilgjengelige for hvem som helst, men inneholder mange muntlige ord og uttrykk og unormert stavemåte.

I tillegg til sortering i teksttyper og sjangre, har vi merket hver enkelt tekst med demografiske variabler: forfatterens alder, kjønn, oppvekststed og fødselsår, der det har latt seg gjøre. Tekstene blir videre merket med utgivelsesår og utgiver/forlag. På den måten er det enkelt å lage subkorpus for undersøkelse av leksikalske egenskaper ved de enkelte ordene hos

forskjellige typer forfattere eller utgivere, eller når de ble skrevet. Leder for prosjektet er professor Ruth Vatvedt Fjeld, godt assistert av avdelingsingeniør Rune Lain Knudsen. Slike leksikalske ressurser er uunnværlige i moderne ordbokarbeid. For det første er det viktig at lemmaseleksjonen støtter seg på frekvens i et godt balansert korpus, og ikke på leksikografenes skjønn eller tilfeldige ordsamlinger. Det gjelder både for å ta inn nye lemmer og for eventuelt å slette utdaterte ord – eller i stedet for å slette dem, å merke dem som sjeldne, lite brukt eller liknende. Hvilke ord man skal velge ut og bestemme og beskrive i en ordbok, er helt avhengig av hvilken type ordbok man vil lage. Ofte tenker man på allmennspråklige ordbøker når norske ordbøker omtales, men det er kanskje vel så stort behov for fagspråklige ordbøker (både for fagfolk og for allmennheten), skoleordbøker, barneordbøker, en- eller flerspråklige ordbøker, bare for å nevne noen. Om man har avgjort hvilken målgruppe en ordbok lages for, kan vi med et korpus som LBK2013 velge ut tekster som passer til formålet og velge oppslagsord ut fra det. Skal det lages en skoleordbok, er det viktig at lærebøker for nettopp det skoleslaget utgjør hovedstammen i korpuset. Med enkle metoder trekker man ut alfabetiserte ordlister av det utvalgte materialet, samt antall forekomster i korpuset. LBK2013 sorterer også ordene automatisk etter ordklasse. Da er man alt langt på vei for å velge oppslagsord for den ordboka man vil lage.

En ordbok er ikke bare en alfabetisk liste over visse ord eller ordtyper. Vi må skille mellom simplex-former og sammensatte former av oppslagsord – vanligvis tar man med sammensatte former bare der betydningen er leksikalisert, dvs. at sammensetningen har fått en betydning som ikke kan leses ut av betydningen til simplexformene som den utgjøres av. Videre er det vanlig å dokumentere flerordsuttrykk, det kan være typiske kollokasjoner (dvs. ord som ofte brukes sammen, som *begå + forbrytelse*), eller idiommer (uttrykk som er relativt uforklarlige, som *skjegget i postkassa*). Flerordsuttrykk kan dokumenteres i spesialordbøker eller som sublemmer i vanlige ordbøker, noe som er det vanligste i norsk (ofte uten at de er skilt fra vanlige eksempler). Statistisk analyse av samvarierende ord kan gi vel så god dekning av flerordslemmer og kollokasjoner som intuitiv innsamling av eksempler og belegg. En kombinasjon av begge metoder er som regel det beste.

Videre kan et korpus vise veldig enkelt og tydelig hvilke bøyingsformer som brukes, og vi kan også finne ut om det er bruksforskjeller mellom forskjellige aldersgrupper, f.eks. etter kjønn eller hvor de bor. Også for språknormering er statistiske undersøkelser av balanserte korpus svært viktig, f.eks. kan en liten og grei undersøkelse vise i hvilken grad hunkjønnsformer fortsatt er i bruk i norsk, og i hvilke grupper av befolkningen. Med samme metode kan vi lett vise når et nytt ord kom inn i språket eller hvor ofte det ble brukt de enkelte år eller perioder man deler korpuset inn i, hvilke andre ord det danner sammensetninger med eller hvilke ord det brukes sammen med. Dette ble illustrert på seminaret med ordet *muslim* og hvordan bruken av det har økt fra 1985 til det i dag er et relativt vanlig ord som inngår i mange slags sammensetninger. Slik er ordundersøkelser også god dokumentasjon på endringer i det samfunnet språket brukes i.

Moderne leksikografi er ikke bare en liste over et språks ordinventar og bøyingsmuligheter for det. Man gir også en oversikt over hvilke ord som faktisk brukes, og hvordan de brukes. Moderne leksikografiske ressurser blir stadig mer sentrale som datagrunnlag også i annen språkforskning, og et godt korpus kan være premissleverandør for mange viktige forskningsprosjekt.

Et større leksikografisk bokmålsprosjekt ble presentert på slutten av foredraget: BRO-prosjektet, som er en forkorting for Bokmålets og riksmålets ordbokverk. Universitetet i Oslo

og Det Norske Akademi for Sprog og Litteratur inngikk en avtale i august 2011 om å samarbeide om å dokumentere hele bokmålets ordforråd og bøyningsverk, med betydningsdefinisjoner, eksempler og domenemarkering, samt kollokasjoner og andre kombinasjonsmuligheter. Dette ordbokverket skal utgis som en digital ordbok og gjøres fritt tilgjengelig på Internett.

Utgangspunktet for BRO er for det første Norsk Riksmålsordbok og den moderniserte varianten NAOB med omfattende dokumentasjon av ordenes betydning og bruk og belegg i litteraturen, for det andre Norsk ordbank, som er en omfattende lemmaliste over bokmålets ordforråd og fullstendig formverk for den offisielle bokmålsnormen. I tillegg skal det utarbeides en statistisk oversikt over kollokasjoner og faste uttrykk med LBK2013 som materiale, og korpuset blir også utnyttet for å gi eksempler på bruk av ordene i forskjellige teksttyper og sjangre. Det er også planen å dokumentere frekvens av de forskjellige formvalgene som brukerne gjør ut fra statistisk opptelling i LBK2013.

De forskjellige komponentene i BRO skal lenkes sammen for en felles digital visning slik at de framstår som ett felles og omfattende ordbokverk med en omfattende og nøyaktig dokumentasjon av ordinventar, og informasjonstyper knyttet til hvert enkelt ord.

BRO-prosjektet vil bli en moderne og utfyllende dokumentasjon av det moderne bokmålet og er gjennomførbart på kort tid og med moderate kostnader, siden de fleste ressursene allerede er utarbeidet og langt på vei ferdig bearbeidet som leksikografisk materiale. Grensesnittet skal ta hensyn til alle de sammenlenkede ressursene fra ILN og DNASL.

Korpusleksikografi i Norge: fra mønsterkorpus til monsterkorpus

Gisle Andersen

Norges handelshøgskole

En viktig forutsetning for effektivt ordbokarbeid og annen språknormering er å ha tilgang til tekstkorpus. Disse bør helst være store og kontinuerlig oppdaterte, slik at de utgjør best mulig grunnlagsmateriale for å følge utviklingen av ordtilfanget og språkutviklingen mer generelt. I Norge har vi flere tiårs erfaring med å utvikle slike korpus, men et problem er at tradisjonelle korpus fort blir utdaterte. I vår digitale tidsalder er det gode muligheter for å utnytte tekster som er tilgjengelige på Internett eller i tekstarkiver som grunnlag for språkbobservasjon. I denne artikkelen vil jeg kort gjøre rede for aktuelle korpus, med hovedvekt på bokmål, og beskrive hvordan jeg oppfatter at vi kan bedre utnytte eksisterende ressurser og teknologi for å sikre bedre tilgang til pålitelige data om norsk språk.

Når man tar i bruk korpuslingvistisk metode i arbeidet med ordbøker, kan det kalles korpusbasert leksikografi eller korpusleksikografi. Sammensetning og utvikling av tekstkorpus avgjøres av hva slags formål som de er tenkt brukt til. Foruten ordbokarbeid kan det dreie seg om språkvitenskapelig forskning, språkteknologisk produktutvikling eller pedagogiske formål. De korpus som er utviklet for norsk, er derfor ulike mht. omfang, sammensetning av teksttyper, datering og grad av annotering av grammatisk informasjon og metainformasjon om teksttype, forfatter osv. Prinsippet bak korpusleksikografien er at man bruker korpuset til å skaffe belegg for ord og uttrykk som er aktuelle for oppføring i ordbøker. Dette omfatter både observasjon av selve ordtilfanget og leksikalsk kunnskap om enkeltord. Korpuset brukes til å kartlegge ordformers frekvens og spredning i korpuset (altså om et ord forekommer i mange ulike tekstkategorier og er brukt av flere forfattere, eller om det har en snever utbredelse), hvilke nyord som forekommer, og om det kan være grunnlag for å ta ut enkelte ord (arkaismer). Spesifikke undersøkelser behøves gjerne for å kartlegge ulike egenskaper ved enkeltord, som skrivemåte, morfologiske egenskaper som kjønn, bøyingsformer, avledninger og sammensetninger (produktivitet), og dessuten ulike betydninger av ordet, ordklasse, bruk i vanlige ordforbindelser osv. Selv om store tekstkorpus er uvurderlige i et slikt kartleggingsarbeid, kan det være grunn til å minne om at det at et bestemt ord eller en ordform forekommer i korpuset, ikke nødvendigvis betyr at det skal inngå i en ordbok. De fleste nyord er flyktige sammensetninger som ikke har «nubbesjans» til å havne i ordboken, enten fordi de er knyttet til en altfor spesifikk brukskontekst, eller fordi de har en betydning som enkelt lar seg avlede ut fra enkeltleddenes betydning. Men også når det gjelder skrivemåte og bøyingsmorfologi, er det slik at dokumenterte korpusbelegg ikke nødvendigvis innebærer at formen skal normeres, for eksempel hvis denne strider mot vedtatte ortografiske prinsipper. Dermed vil korpusbasert arbeid aldri kunne erstatte de manuelle analyser og kvalitative vurderinger som leksikografen må foreta.

I vår tid kan det være fristende å tenke seg at det er tilstrekkelig å google et ord for å finne ut om det er i bruk eller ikke. Dette er for de aller fleste formål ikke godt nok, og innen korpuslingvistikken er det for lengst slått fast at bruken av Google for språklige dokumentasjonsformål er utilstrekkelig og innebærer en rekke svakheter. Resultatene av slike søk gir for lite brukskontekst, for få treff vises om gangen, visningsrekkefølgen er ikke manipulerbar og er dessuten styrt av kommersielle prinsipper, statistikken er upålitelig, og

dessuten er populasjonen (materialets totale størrelse) ukjent. Bruken av godt annoterte tekstkorpus er derfor en forutsetning for å dokumentere faktisk språkbruk. Leksikografien krever data som er relevante for leksikografiformål, dvs. data som er tilstrekkelig omfangsrike, oppdaterte og gode. Men det er dyrt å bygge korpus, og vi må prioritere å utvikle ressurser som er realistisk å fremskaffe og gjøre tilgjengelig innenfor gjeldende rammer for finansiering. Spørsmålet blir da: hvordan kan vi best mulig utnytte eksisterende ressurser som korpus, korpusverktøy, kompetanse, store tekstarkiver og datateknologi i leksikografien?

Blant eksisterende korpus finner vi Oslo-korpuset av taggede norske tekster og Leksikografisk bokmålskorpus (jf. Fjelds innlegg), og dessuten Språkbankens gullkorpus og INESS-korpuset, som begge er under utvikling. Dette er gode ressurser som utvilsomt har en verdi for leksikografien. Slike tradisjonelle korpus er kostbare å utvikle og forholdsvis kompliserte å samle inn, og fordi de er statiske ressurser, er oppdatering mer komplisert. Av den grunn er det etter hvert blitt vanlig å bruke nettbaserte korpus i leksikografien og ellers. De har den fordel at de er langt billigere å utvikle og enklere å samle inn, og de er dynamiske og kan være kontinuerlig oppdaterte. Det kan være verdt å merke seg at leksikografene bak den omfangsrike Oxford English Dictionary nå bruker hovedsakelig et nettbasert korpus, dvs. Oxford English Corpus, som består av 3 mrd. ord samlet inn fra nettet i perioden fra 2000 og frem til i dag. De mest kjente norske nettkorpus er Norsk aviskorpus, som inneholder 1 mrd. ord av tekster fra norske nettaviser, og NoWaC, som har internettekster fra mange ulike kilder. Erfaringene fra prosjektet Norsk aviskorpus har vist at nettbasert korpusinnsamling er enkelt å sette i gang, og det er enkelt å kontrollere teksttilfanget ved å bruke et begrenset utvalg domener. Problemet med denne typen korpus er imidlertid behovet for duplikatkontroll og fjerning av rusk og uønsket tekst.

I tillegg til nettbaserte korpus kan det være aktuelt å bruke andre tekstsamlinger som grunnlag for språklig dokumentasjon. Nasjonalbibliotekets ressurser er svært omfattende, og Bokhylla.no inneholder per i dag 50 000 bøker, aviser og manuskripter fra 1690–2000. Når den er ferdigutviklet i 2017, skal Bokhylla inneholde 250 000 bøker. En annen omfattende ressurs er Retrievers database ATEKST, som inneholder 48 aviser og tekster fra internett og andre medier fra 1945 frem til i dag. Slike tekstarkiver er helt klart relevante for ordbokarbeid. Problemet med disse ressursene per i dag er at de ikke fremstår som søkbare korpus, og man får frem ordforekomster i pdf-formatet, som er uegnet for videre bearbeiding av dataene og atskillig mindre nyttig enn den konkordansevisning som er vanlig i korpus. Arkivene gir i noen grad muligheter for begrenning iht. tekst/emnekategori, år, forfatter osv. men de mangler den mulighet for sofistikerte søk etter språklige kategorier som er vanlig i korpus, slik som å søke etter en lemmaform og dermed finne alle bøyningsformer av et ord, søkbegrensning etter ordklasse eller morfosyntaktiske trekk, eller mulighet for å skille appellativer (*blitz* eller *blits*) fra proprier (*Blitz*).

Redegjørelsen over viser at vi har rikelig med teknologi, dataverktøy og metoder fra korpuslingvistikken og språkteknologien som kan og bør brukes til å forenkle arbeidet med ordbokredigering og språknormering. Vi har dessuten også rikelig med erfaring og kompetente FoU-miljøer som kan ta korpusleksikografien videre. Det er viktig at normeringsforslag i ordbøker og ellers fremstår som velbegrunnet og konsistent. Derfor trenger vi systemer for effektivt å normere alle likelydende ord i sammenheng og også å se bokmål og nynorsk i sammenheng. Dermed vil min viktigste anbefaling være å få til en kobling av korpusmetoder til de store tekstarkivene og dermed få etablert det vi kan kalle for et «monsterkorpus», det vil si et svært omfattende og voksende korpus. Dette dreier seg altså

ikke om å samle inn nytt materiale, men å ta i bruk tekster som allerede er tilgjengelige og arkivert og gjøre det mulig å bruke disse med bedre funksjonalitet for søkbegrensning og visningsmanipulering i tråd med korpuslingvistiske prinsipper og arbeidsmåter. Ideen om «Bokhylla som korpus» går altså ut på å bruke Nasjonalbibliotekets ressurser på en måte som er tilpasset leksikografiformålet, men som også vil være en svært viktig forskningsressurs. Dette vil være det viktigste enkelttiltaket for å lette arbeidet med ordbøker og språknormering og vil gjøre livet enklest mulig for leksikografer, som er opptatt av mest mulig oppdaterte data om faktisk språkbruk.

Kva slags språksamlingar treng Noreg etter 2014?

Oddrun Grønvik
Norsk Ordbok 2014

Dette spørsmålet er lett å svara på. Noreg treng godt utbygde og lett tilgjengelege morsmålssamlingar med høg vitskapleg standard, både for norsk og for dei andre morsmåla som Noreg har politisk ansvar for. Morsmålssamlingane må dekkja språka som heilskap, og syna språka både over tid (diakront) og i dei variantar som finst i samtida (synkront). Språksamlingar med høg vitskapleg standard er grunnlag for vidare forskning, pålitande språkverktøy og informasjonskjelder for folk flest. Utan vitskaplege språksamlingar kan ein ikkje laga dei språkverktøya som moderne språk treng, verktøy som operasjonaliserer offisiell rettskriving, skrivereglar, grammatikk og ordtilfang (ordbøker).

Både morsmålssamlingane og språkverktøy som byggjer på samlingane, treng tillit frå språksamfunnet dei skal tena. Det oppnår ein enklast ved at morsmålssamlingane er i offentleg eige og lett tilgjengelege for flest mogleg. Som norsk språkhistorie tydeleg viser, er openheit og medverknad føresetnader for tillit, og i siste instans for konsensus om resultatata.

Kvifor er dokumentasjon av eigne morsmål ei sentral språkpolitisk oppgåve? Dokumentasjon av morsmål er ein av dei viktigaste måtane å byggja samfunnsmessig infrastruktur på. Dokumentasjonen trengst av di språket – morsmålet – er det fremste provet eit språksamfunn har på at det eig ei sams historie og ei sams omgrepsverd. Språk blir ikkje til over natta. Språk er umåteleg omfattande instrument for utveksling og lagring av informasjon, så elegant oppbygde at småbarn kan tileigna seg det grunnleggjande, medan ingen nokon gong kan gjera krav på å meistra morsmålet sitt heilt til endes. Er språket godt dokumentert, blir det umogleg å sjå bort frå at språksamfunnet finst.

Godt dokumenterte språk gjev grunnlag for skriftlege standardspråk til bruk i opplæring og kommunikasjon. Språksamfunn som manglar dokumenterte standardar for morsmåla sine, får utdanningstilbod og administrasjon på andre folks standardspråk. Dette er nær forhistorie og til dels realitet i Noreg. Dokumentasjon av morsmåla er ein føresetnad for at språksamfunn skal kunna fungera som utdanningssamfunn og delta i alle dei skriftbundne samfunnsfunksjonar som moderne demokrati kviler på. Dette er like sjølvstøtt for morsmåla i Noreg som i andre samfunn.

Spørsmålet om kvifor morsmål skal dokumenterast, og kvifor dokumentasjonen skal vera lett tilgjengeleg for alle, treng ikkje forfølgjast her. Det som derimot må omtalast, er følgjande:

- kva slag morsmålsdokumentasjon trengst for kvart språk?
- kva slag morsmålsdokumentasjon finst for kvart språk?
- kva slags rammer bør dokumentasjonsverksemda ha når det gjeld institusjonalisering, finansiering og administrativ innplassering?

Kva slags morsmålsdokumentasjon trengst for eit språk?

Alle språksamfunn har same behov for språkleg dokumentasjon. Kor langt og korleis denne dokumentasjonen kan oppnåast, er derimot eit emne som må drøftast språk for språk.

Alle språk treng dokumentasjon i form av samlingar som dette:

- tekstsamlingar som syner språket i bruk – skrift og tale (for teiknspråk: teikn og transkripsjon)
- dømesamlingar som grunnlag for grammatisk analyse med lydsystem, bøyning, ordlaging og setningsoppbygging (syntaks)
- indeks over einskildord i standardform til å samordna tilfang som høyrer saman, men har ulik uttale eller skrivemåte
- aamlingar for særspåk (terminologi, namnesamlingar)

Ivar Aasen dekte desse behova for norsk med grammatikkar, ulike slag ordbøker, tekstar med målprøver og anna. Ein liknande innsats har andre lingvistar gjort i møte med udokumenterte språk.

I moderne, elektronisk form bør kvart språk ha

- store tekstkorpus med god søkjefunksjonalitet
- digitale indeksar over oppslagsformer for einskildord (som t.d. Metaordboka for nynorsk)
- fullformsgenerator (t.d. Ordbanken)
- morfosyntaktisk taggar (t.d. Oslo-Bergen-taggar)
- andre verktøy for syntaktisk analyse (t.d. i form av analysesystemet attom ein trebank)
- databasar med termar, namn, og andre særspåksamlingar

Samlingar og verktøy bør vera lett tilgjengelege, lette å bruka, og så samordna som fagleg forsvarleg.

Om ein måler eksisterande dokumentasjonen av dei morsmåla som det offentlege Noreg har ansvar for mot lista ovanfor, ser ein at ingen har full dekning. For bokmål og nynorsk er stoda betre enn for samisk, kvensk, romani og teiknspråk. Det blir viktig å sørgja for at alle morsmåla får jamførbar dokumentasjon, og at dokumentasjonen av kvart morsmål held høg vitenskapleg standard.

Kva slags tilbod?

Språksamlingane for morsmål i Noreg bør vera eit permanent og sjølvstendig tilbod til alle språkbrukarar. Språksamlingane og språkprodukta som byggjer på dei, blir brukte i forskning og språkleg normering, men også på fleire måtar og til fleire føremål enn nokon kan føreseia. Folketeljningane blir til jamføring brukte i forskingsprosjekt og offentlege rapportar, men også til ei mengd føremål som ingen har full oversikt over.

Eit permanent og sjølvstendig tilbod om tilgjenge til språksamlingane i Noreg inneber at samlingane må liggja føre på Internett i elektronisk form. Dei må vera lette å finna og søkja i. Dei bør vera gratistenester der det offentlege garanterer for kvalitet og tilgjenge, liksom Artsdatabanken og samlingane til Statistisk sentralbyrå. Fritt tilgjenge til språksamlingane vil styrkja interessa for og engasjementet i morsmåla i Noreg, og fremja kunnskap og bruk. Elektroniske språksamlingar og språkverktøy på Internett som er lette å bruka og gratis i bruk, er internasjonalt tilgjengelege, og gjev noregsreklame verda over.

Rammer for dei vitenskaplege morsmålssamlingane i 2013

Dei fleste vitenskaplege morsmålssamlingane som dokumenterer nyare språkstadium, er i dag lokaliserte ved norske universitet og høgskular. Dels er dei blitt til gjennom lokal forskarverksemd, ofte gjennom fleire generasjonar (Hannaas-samlinga ved UiB; Målføresynopsisen, Norsk Dialektatlas ved UiO). Dels er dei deponerte ved institusjonen. Det gjeld såleis samlingane som gjekk inn i Norsk leksikografisk institutt (1972) og Institutt for namnegransking (1978), båe ved UiO.

Vitenskapleg, systematisk morsmålsdokumentasjon ved Universitetet i Oslo byrja på 1880-talet. På 1900-talet var språkleg dokumentasjon dels ei permanent verksemd (Norsk målførearkiv, UiO), dels knytt til ulike prosjekt med målsetjing om å byggja permanente samlingar (namnearkiva, ordbokverka). På 2000-talet er viktige samlingar komne til som har dei eldre samlingane til jamføringsgrunnlag, såleis Nordisk dialektkorpus ved UiO. Fleire forskarmiljø har bygd opp store tekstkorpus til vitenskapleg bruk. Universitetet i Tromsø stør opp om mykje av det som er gjort innanfor dokumentasjon av samisk og kvensk.

Dokumentasjonen av morsmåla i Noreg før 1811 er spreidd på fleire land og institusjonar, særleg er det mykje ved Det Kongelige Bibliotek i Danmark. Å laga ei elektronisk samling av eldre kjelder for morsmåla i Noreg har lenge vore eit ønskemål som det ikkje har vore mogleg å finansiera. Dette gjeld først og fremst mellomnorsk og tidleg nynorsk.

Forsking genererer samlingar. Nye kunnskapsbehov vil føra med seg fleire språksamlingar, som t.d. korpus av elevtekstar, men òg mindre samlingar knytte til individuelle prosjekt. Det vituge er då å ha eit system for ivaretaking og gjenbruk – noko som manglar i dag.

Rammene for dei vitenskaplege morsmålssamlingane etter 2014

Framtida er uviss for alle morsmålssamlingane ved universitet og høgskolar. Humanistisk fakultet ved Universitetet i Oslo reknar språksamlingane som ikkje finansierte frå utlaupet av 2014. Då fell driftsmidlane bort, og stillingane til det noverande vitenskaplege personalet vil ikkje bli fylt ved vakansar. Det inneber at språksamlingane ved UiO er ubemanna, og dermed nedlagde, innan 4–5 år. Andre universitet og høgskolar har ikkje sett sluttdato, men heller ikkje lagt planar for vidareføring.

Vitenskaplege samlingar forvitrar om dei ikkje blir brukte. Papirsamlingar kan oppbevarast lenge, endå om brukskompetansen går tapt når det vitenskaplege personalet forsvinn. Verre er det med samlingar i andre medium, til dømes ikkje-digitaliserte lydband. Også elektroniske samlingar krev vedlikehald. Både programvare og maskinvare blir forelda. Elektroniske samlingar og produkt som i dag er førsterangs, kan bli ubrukelege på få år om fagkompetansen til vidareføring manglar.

Spørsmålet om kva som skal skje med dei vitenskaplege språksamlingane ved universitet og høgskolar, har lege på vent i Kulturdepartementet sidan 2008, og den varsla politiske handsaminga er ikkje komen enno².

² Det er ikkje kome noka politisk utspel om organiseringa av språksamlingane etter St.meld. nr. 35 (2007–2008)

Ansvar for språksamlingane er i praksis delt mellom to departement. Dette vart ikkje tematisert då Vitskapsavdelinga i 1991 vart overført frå Kulturdepartementet til Kunnskapsdepartementet. Kulturdepartementet har i dag det overgripande ansvaret for språk som politikkområde, og finansierer dei største språkprosjekta. Men Kunnskapsdepartementet har sidan 1991 ansvar for universitet og høgskolar, og dermed også for insentivsystema som avgjer kva haldning universitet og høgskolar syner til språksamlingar under eige tak. Dei vitskapelege språksamlingane ligg i dag i grenseområdet mellom verkefeltet for dei to departementa, og kan i verste fall bli kasteball mellom dei, eller usynleg for både.

Då NLI og seinare IN vart skipa ved UiO på 1970-talet, var tanken å sikra langsiktige trygge rammer for både morsmålssamlingane med tilhøyrande vitskapeleg verksemd. Det er opplagde føremonar i å knyta forskingsverksemda ved morsmålssamlingane til anna språk- og kulturforskning, og organisasjonsforma har på dette punktet innfridd føremålet. Derimot var det neppe påtenkt at resursane som følgde med desse språksamlingane, skulle bli omdisponerte³ ved UiO, men det er skjedd.

Ei ny administrativ løysing for morsmålssamlingane bør derfor halda på føremonen med å vera del av eit større, aktivt forskingsmiljø, men vera mindre sårbar for institusjonell omorganisering. Det er også rimeleg at samfunnsansvaret til morsmålssamlingane blir tydeleg uttrykt, og at samlingane får ei leiing med tilsvarende handlingsrom.

³ NLI og IN hadde i 1991 19 vitskapelege og 3,5 administrative faste stillingar. Av desse er 5 att, resten er omdisponert av UiO i samband med naturleg avgang. I denne perioden er Målførearkivet blitt nedlagt på same måte.

Ressursar og kompetanse i Norsk Ordbok 2014-prosjektet

Åse Wetås
Norsk Ordbok 2014

Innleiing

Norsk Ordbok er ei vitskapleg ordbok i 12 band som dekkjer det norske folkemålet (alle dei norske dialektane) og det nynorske skriftspråket. I ferdig stand vil det trykte verket omfatta meir enn 300 000 ordartiklar. Artiklane i Norsk Ordbok er dessutan fritt tilgjengelege for ålmenta i ei nettutgåve lansert i 2012. Nettutgåva omfattar i dag alle artiklar som er produserte elektronisk i Norsk Ordbok 2014-prosjektet (alfabetstrekket frå og med *i*–). I tillegg til desse to konkrete forskingsprodukta omfattar ressursane i NO 2014 ein solid og funksjonell redigeringsapplikasjon som er lenkja saman med digitalisert grunnlagsmateriale i eit system av relasjonsdatabasar, dei omfattar eit nynorsk tekstkorpus som dekkjer perioden 1866–2013 og dei omfattar sjølvstøtt den breie leksikografiske kompetansen til forskarkollegiet som står bak utgjevinga.

Kort om verket Norsk Ordbok og prosjektet NO 2014

Norsk Ordbok gjev ei vitskapleg framstilling av det norske folkemålet og det nynorske skriftspråket. Arbeidet med ordboka blei starta i 1930 av Det Norske Samlaget, men ordboka og dei store samlingane med grunnlagsmateriale blei i 1972 lagde inn under Universitetet i Oslo, som eit resultat av konklusjonane frå den politiske handsaminga av Vogt-komiteens innstilling i språksaka frå 1966. Med innlemminga i UiO blei ordbokverket del av eit nyoppretta Norsk leksikografisk institutt.

Fram mot årtusenskiftet blei det teke initiativ til å ferdigstilla Norsk Ordbok i 12 band til 200-årsfeiringa av den norske Grunnlova i 2014. Prosjektorganisasjonen Norsk Ordbok 2014 blei etablert i 2002 etter ein avtale mellom UiO og Kulturdepartementet. Ein ambisiøs og forpliktande utgjevingsavtale blei sett opp, og det blei laga ein redaksjonell opptrappingsplan som skulle sikra at det var mogleg å nå dei oppsette måla. Band 1-4 av verket hadde komme ut i perioden 1966 til 2002. I prosjektperioden er band 5 t.o.m. band 10 av verket gjevne ut, og redaksjonen leverer sommaren 2013 band 11 til forlaget for prenting og utgjeving.

Norsk Ordbok i ei ny tid

På 1990-talet blei vesentlege delar av grunnlagsmaterialet til Norsk Ordbok digitalisert gjennom Dokumentasjonsprosjektet. I den tidlege fasen av NO 2014-prosjektet utvikla Eining for digital dokumentasjon (EDD) ved UiO ei digital redaksjonell produksjonsplattform i nært samarbeid med prosjektorganisasjonen. Redaksjonen har eit laupande drifts- og utviklingssamarbeid med EDD, og i prosjektperioden har redigeringsapplikasjonen blitt oppgradert med viktig tilleggfunksjonalitet. Heile applikasjonen er nyskriven i perioden 2010-2013.

Norsk Ordbok er eit vitskapleg ordbokverk. Dei digitale verktøya som er utvikla for å redigera og publisera verket, sikrar tett integrering av kjeldemateriale og sluttprodukt. Den elektroniske indeksen ”Metaordboka” (MO) dannar kjernen i applikasjonen og ligg som eit nav i systemet av relasjonsdatabasar. I Metaordboka er alt kjeldemateriale frå dei ulike

materialdatabasane samsortert under normerte oppslagsord. Indeksen har om lag 600 000 oppslagsnoder, og desse kan ha kopla til seg alt frå eitt til fleire tusen belegg frå kjeldesamlingane. I tillegg til innsamla leksikografisk materiale, digitaliserte ordsamlingar og eldre kanoniske ordbøker omfattar grunnlagsmaterialet til Norsk Ordbok også Det nynorske tekstkorpuset. Korpuset er bygd opp i prosjektperioden og inneheld i 2013 om lag 90 mill. ord (fordelte på ulike sjangrar og tidsperiodar). Det innsamla materialet og tekstkorpuset utfyller kvarande på vesentlege måtar, og korpuset var òg ei viktig kjelde i Språkrådets arbeid med 2012-normalen for nynorsk. Både korpuset og det store innsamla grunnlagsmaterialet er fritt tilgjengleg på NO 2014s nettside.

Utdjupande om verktøya

Norsk Ordbok er ei vitskapleg ordbok. Redigeringsverktøya våre omfattar ein redigeringskomponent som produserer tekst i xml, og eit sett relasjonsdatabasar og program som gjer redigeringsapplikasjonen til eit integrert anlegg for produksjon, analyse, kontroll og dokumentasjon. Alle opplysningar i verket skal kunna etterprøvast på ein enkel og direkte måte, og alle ordartiklar i verket har difor ei elektronisk kopling til grunnlagsmaterialet gjennom den nemnde MO-indeksen (jf pkt. 3).

Ut frå talet på registrerte ordbelegg i MO-indeksen reknar applikasjonen ut tilrådd omfang for kvar einskild ordartikkel i verket. Denne funksjonaliteten er vesentleg for at organisasjonen både skal kunna sikra alfabetprogresjonen og samstundes halda seg til dei vedtekne produksjonsrammene.

Redigeringsverktøya kommuniserer òg direkte med databasar over relevante metadata for verket. Dette sikrar korrekt lenking og korrekte referansar til bibliografiske opplysningar, bruksmerkingar, heimfestingar osv. Den kvantitative produksjonen kan målast laupande i rapporteringsdelen av applikasjonen, og dette sikrar full kontroll både med framdrift og med alfabetprogresjon. Denne funksjonaliteten gjer at oppdragsgjevarane våre når som helst kan få informasjon om kvar produksjonen står i høve til dei oppsette planane.

Verktøysettet omfattar dessutan eit system for digital arbeidsflyt som sikrar at alle artiklar går gjennom ei intern kvalitetssikring i form av ei rekkje kollegalesingsinstansar, fram mot endeleg produksjonsstatus "Ferdig". Dei ferdige ordartiklane er lagra i xml med backupversjonar og blir publiserte både elektronisk (html) og i trykt versjon (pdf).

I og med NO 2014-prosjektet er store mengder språkdokumentasjon digitalisert og tilgjengeleg for ålmenta over nettet. Dei leksikografiske forskingsverktøya som er utvikla i prosjektperioden, møter behovet for å sikra etterprøvarheit gjennom enkelt tilgjenge til dokumentasjons-ressursane som ligg bak verket, dei gjer det mogleg å dimensjonera arbeidet, dei sikrar akribi og arbeidsflyt, og dei gjev naudsynt kontroll med framdrifta.

IT-ressursane er knytte til eit stort og fagleg kompetent utviklingsmiljø ved UiO. Dette er viktig for stabilitet og driftssikkerheit og ikkje minst til høve for kontinuerleg vidareutvikling av ressursane.

Forskarstaben

Alle redaktørane i NO 2014 er forskarar med doktorgrad eller mastergrad i språkvitskap. Fire er professorar. Alle forskarane har fått eitt års redaksjonsintern opplæring, i form av eit redaksjonelt kursopplegg i kombinasjon med individuell rettleiing. Artikkelproduksjonen er meritterande vitskapleg publisering for den einskilde forskaren. Dette fører til at

arbeidsplassane er attraktive for språkvtitarar med forskingsambisjonar, det sikrar stabilitet i staben, og det sikrar høg kvalitet på sluttproduktet.

Nettutgåva av ordbokverket

Dei digitale verktøya er utforma slik at prosjektet har kunna utvikla ei fritt tilgjengeleg nettutgåva utan vesentlege meirkostnader. Nettutgåva omfattar dei delane av verket som er produserte på digital plattform (frå og med alfabetbokstaven *i*). Grunnlagsmaterialet og sluttprodukta er tilgjengelege til både forskingsføremål og til offentleg språkdyrking og språkstyrking. Dei er dessutan gratis nettilgjengelege for alle som er interesserte i norsk språk.

Andre resultat

Norsk Ordboks redigeringsapplikasjon har eit maksimumsformat til språkdokumentasjon, men kan med hell også brukast på andre typar ordbøker. *Nynorskordboka* og *Bokmålsordboka* er for tida under revisjon, og dette arbeidet blir gjort i NO-applikasjonen. Det har lagt til rette for å etablere unike koplingar mellom ordartiklar i dei to verka og relevant bøyingsinformasjon i Norsk ordbank, til glede for alle brukarar av handordbøkene på nett. Redigeringa av handordbøkene i dette applikasjonsformatet aukar kvaliteten på verka, det aukar tempoet i revisjonsarbeidet, og det gjev eit stringent system som er enkelt å vedlikehalda. I tillegg representerer det god gjenbruksøkonomi.

Framtida

Dei eldste banda av Norsk Ordbok (alfabetstrekket *a-h*) er ikkje digitalt tilgjengelege. Det trengst ein innhaldsrevisjon og ei digital fullintegrering av dette alfabetstrekket for at databasen skal bli komplett og kunna brukast til andre føremål i framtida.

NO 2014-organisasjonen gjorde i 2006–2007 ein pilotstudie av arbeidsmengda knytt til revisjon og elektronisk integrering av materiale frå dei eldste banda av Norsk Ordbok. Basert på denne pilotstudien er det sett opp ein prosjektplan for eit slikt arbeid, som vil gjera e-utgåva av verket fullstendig. Revisjon av dei eldste banda av verket blei varsla som framtidig arbeidsoppgåve alt i forprosjektrapporten ”Norsk Ordbok 2014”, som Steinulf Tunesvik utarbeidde for UiO i år 2000. Tunesvik konkluderte der med at dette var ei oppgåve som ville vera så arbeidskrevjande at ho måtte definerast utanfor 2014-prosjektet.

I St.meld. nr. 35 (2007–08) heiter det om ein framtidig dokumentasjon av bokmålet at ”departementet legg til grunn at eit slikt prosjekt i regi av Universitetet i Oslo vil ha stor nytte av kompetanse og ressursar som er bygde opp i samband med nynorskprosjektet Norsk Ordbok 2014” (kap. 8.3.6). Dersom planen er å gje bokmålet ein tilsvarande ambisiøs, vitskapleg dokumentasjon som den nynorsken har fått gjennom 2014-prosjektet, vil det også for den målforma kunna hentast ut langsiktige tilleggsvinstar i form av ålment tilgjengelege og multifunksjonelle dokumentasjonsressursar av same typen som dei som nå er etablerte for nynorsk. NO 2014 har i prosjektperioden bygd opp kompetanse og ressursar som vil vera svært nyttige i eit slikt arbeid.

Avslutning

Det er fagleg sett fullt mogleg å gjennomføra eit prosjekt for digitalisering og revisjon av dei eldste NO-banda side om side med eit prosjekt for vitskapleg dokumentasjon av bokmål. Det vesentlege nå er å få ei rask avklaring på korleis kompetansen og ressursane som er utvikla i NO 2014-prosjektet i tida etter 2002, kan brukast for framtida.

Om tilstanden og behovene for ordbøker for innvandrere

Olaf Husby

Institutt for språk- og kommunikasjonsstudier
Norges teknisk-naturvitenskapelige universitet (NTNU)

Per 1.1.2013 var det registrert innvandrere fra 245 land i Norge. Det finnes ingen oversikt over hvor mange språk som er representert i landet. Grunnskolens Informasjonssystem⁴ oppgir at det i 2012/13 gis undervisning i 109 morsmål i grunnskolen. Det er rimelig å anta at det faktiske antallet språk i Norge er langt større.

Ordbøker er en viktig og nødvendig ressurs når en skal lære norsk og senere bruke språket aktivt. Det er derfor naturlig å spørre om i hvilken grad det finnes ordbøker for de ulike innvandrerspråkene. De mest solgte ordbøkene i Norge er ikke kommet til som et resultat av innvandringen de siste 50 årene. Hoveddelene er ordbøker som reflekterer fremmedspråksundervisning i skolen og språk som brukes i land som Norge har hatt nær kontakt med gjennom lang tid. Innvandringen siden 1960-tallet har imidlertid ført til at en rekke ordbøker knyttet til andre språk er utviklet.

Ved NTNUs bibliotek finnes det en egen samling av ordbøker som dekker cirka 40 språk. De fleste ordbøkene inneholder oversettelser til/fra norsk, men også andre ordbøker til/fra dansk/svensk er stilt opp. I bibliotekets samling ellers finnes det ordbøker tilknyttet mer fjerne språk, gjerne med oversettelser til engelsk, fransk eller tysk.

I den nevnte samlingen finnes det enveis ordbøker *Norsk/SpråkX* og *SpråkX/Norsk*, men ikke alltid som par. Det vil si at en kan finne oversettelser den ene veien, men ikke den andre. I tillegg finnes toveis ordbøker i ett bind på en del språk. Utover dette finnes fagordbøker, ordlister til lærebøker og terminologilister.

Et generelt og tydelig trekk ved mange av ordbøkene er at de er rettet inn mot norske brukere. Det vil si at de kan gi grunnleggende informasjon om oppslagsordene på det fremmede språket, det vil si opplysninger om ordklasse, bøyningsmønster og kanskje uttale, mens tilsvarende informasjon om norske ord kan være fraværende. En relativt rask gjennomgang av de oppstilte ordbøkene viste at av toveis ordbøker manglet en av tre *Norsk/SpråkX*-ordbøker morfologisk informasjon om det fremmede språket, mens to av tre *SpråkX/Norsk*-ordbøker mangler tilsvarende informasjon om norsk. Noen få ordbøker har denne informasjonen begge veier. Ytterst få ordbøker angir uttale av norske ord. Den raske gjennomgangen viste et sett med ordbøker som pekte seg positivt ut, nemlig *Ny stor norsk-russisk ordbok* (Berkov 2010) og *Ny stor russisk-norsk ordbok* (Berkov mfl. 2006) Her finner en i begge ordbøker informasjon om ordklasse, bøyning og uttale, både språklyder og prosodi.

Omfang

Ordbøkene er av varierende størrelse. Mens de største inneholder rundt 80 000 ord (*Stor norsk-kinesisk ordbok*, Rekkedal mfl. 2005), inneholder de minste bruksordbøkene rundt 5000

⁴ <https://gsi.udir.no>

ord. Ordlistene til innføringsbøker kan inneholde så lite som 1500 ord. Lexin-ordbøkene, som ligger på nettet, oppgis å ha cirka 17 000 ord.

Målform

De fleste ordbøker er på bokmål. Når det gjelder angivelse av bøyingsformer av substantiv og verb, er det generelle inntrykket at mer moderate bokmålsformer er dominerende. Substantiv som kan være både hankjønn og hunkjønn, presenteres oftest som hankjønn, og svake verb i *kaste*-klassen får oftest oppgitt “-et” som preteritumsform. Et unntak her er *Norsk–engelsk ordbok* (Haugen 1995). Boka inneholder 60 000 oppslagsord på bokmål og nynorsk med oversettelser til amerikansk-engelsk. Den viser uttale og gir grammatiske opplysninger om ordklasse og alternative bøyingsendelser. Boka er ikke i salg lenger.

Andre typer ordbøker

Det kan neppe være overraskende at det er det behov for flere ordbøker innenfor andrespråksfeltet. Brukerne her er av to kategorier. For det første finner en minoritetsspråklige som har behov for ordbøker mellom norsk og sine morsmål. Det er åpenbart at den norske delen av disse ordbøkene må inneholde informasjon om ordklasser og bøyning. Hvilke språk ordbøkene bør forefinnes på, kan en til en viss grad avlese av innvandrersjakkene, men i den grad at innvandrerne kommer fra flerspråklige land, må behovet avklares. Slike ordbøker vil også være av nytte for lærere. Ordbøkens bruksverdi ville øke dersom de også kunne presentere sammenlignende beskrivelser mellom norsk og elevenes morsmål.

Forskeren og lærebokforfattere som er tilknyttet andrespråksfeltet, vil kunne ha spesielle behov for ordbøker. Disse kan gjerne være i form av databaser der en kan skreddersy tilbudet. En kan her nevne tilgang til ulike tekstsamlinger, oversikter over fagterminologi og baklengsordbøker. Det er også ønskelig med ressurser slik at en raskt kan gjøre analyser av lærebøker og ulike tekster.

Uttale

Utenlandske studenter og kolleger bemerker ofte at de finner fraværet av uttaleangivelser for norsk i ordbøkene bemerkelsesverdig. De har heller ikke sympati for det faktum at de ikke får tilgang til uttaleordbøker ut fra forhold knyttet til norsk historie og språkpolitikk.

I *Engelsk blå ordbok* (Skaug mfl. 2010) finner at ”ghost” er et substantiv som uttales /gəʊst⁵, og at trykkplasseringen i ghostwriter” (*subst.*) er / gəʊst r tə/, finner en ingen motsvarende informasjon om det norske ordet ”spøkelse” annet enn at det er et substantiv, men er det ”en/ei/et spøkelse”? Og hva er uttalen? Og hvilket tonelag har ordet?

Trykk på første stavelse og lang vokal?	/ ¹ spø:kelse/, / ² spø:kelse/
Trykk på første stavelse og kort vokal?	/ ¹ spøkelse/, / ² spøkelse/
Trykk på andre stavelse og lang vokal?	/spø ¹ ke:lse/, /spø ² ke:lse/
Trykk på andre stavelse og kort vokal?	/spø ¹ kelse/, /spø ² kelse/
Trykk på tredje stavelse og lang vokal?	/spøkel ¹ se:/

De nevner også at de må konsultere tekst- eller grammatikkbøker for å avklare om flertallsformen er ”spøkelse” (jf. et hus – to hus”) eller om den er ”spøkelser”. Denne informasjonen kunne ha vært gitt i regelform som i Bokmålsordboka (n2) eller skrevet ut som endelser ”spøkelse, n; -et, -r, -ne”.

Det er naturlig, men kanskje ikke fullt rettferdig, å sammenligne norske ordbøker med store engelske ordbøker. Det er relevant ut fra det store antallet språk i Norge, men det er ikke rettferdig ut fra at norsk som andrespråk har en relativt kort historie sammenlignet med engelsk. De store engelske ordbøkene inneholder ekvivalenter, ordforklaringer, ordklassetilhørighet, bøyningsmønster, uttale, semantisk informasjon og i noen tilfeller også encyklopedisk informasjon.

De nettbaserte Lexin-ordbøkene⁶ inneholder informasjon om morfologi og uttale. Uttaleinformasjonen gis i form av transkripsjon og syntetisk tale. Transkripsjonen er en blanding mellom ortografi og IPA. Et ord som ”budsjett” skrives ut som /budsjet̥/. Her er plasseringen av ordtrykket markert gjennom understreking, et uttrykk som normalt angir lang vokal i ordlister som følger lærebøker. I tillegg er digrafen <sj> brukt for å angi sje-lyden.⁷ Transkripsjonen kan dermed raskt lede til uttalen /bud sje:t/ istedenfor /bʊdʃet/. Lydeksemplet vil riktignok være veiledende for uttalen, men tekst-til-tale-syntesen er ikke pålitelig. Et tilfeldig søk i Nynorskordboka til Lexin gav uttalen /le s :r/ for ”(ein) lesar”, et ord der en skulle forvente /le:s r/. Ordbokas egen angivelse er /le:s r/, en form som altså står i strid med den faktiske uttalen. Som en ser, er angivelse av tonelag utelatt i Lexin.

Notasjon

Det er ulike konvensjoner som ligger til grunn for transkripsjon av uttale. En kan finne ord skrevet ut i ordinær form supplert med apostrof eller understreking som skal markere trykkplassering og lang vokal. I andre tilfeller finner en at enkelte lydskrifttegn er introdusert, men langt fra alle, slik at digrafer opptrer. Ordbøker med lydskrift etter fonetiske alfabet, for eksempel IPA, er nesten fraværende. Den eneste større utgaven er Vanvik (1985), men den er ikke i salg.

Kognatordbøker

Med utgangspunkt i baklengsordlister kunne det utvikles kognatordbøker, dvs. ordbøker over beslektede ord, som kunne vise overlappende vokabular mellom vestlige språk. Slike ordbøker er utviklet mellom engelsk og andre språk (Nash 1997). Jeg har gjennomført noen usystematiske analyser av kognater i norsk-spansk, og funnene viser at det kan være så mange som 10 000 overlappinger mellom de to språkene. Overlappingen kategoriseres da etter fire nærhetsrelasjoner, og falske venner er også inkludert⁸.

⁶ <http://decentius.hit.uib.no/lexin.html>

⁷ En kan også notere bruken av <u> istedenfor IPA-tegnet <ʊ>, men her er ordbøkene konsekvente idet tegnene <o> og <å> representerer o-lyden og å-lyden istedenfor IPA-tegnene <u>, <o>.

⁸ www.cognates.org: ”statistical data [...] support the assertion that cognates represent at least 25% of the unique English written words met by Romance language speakers, and vice versa.”

Antonymordbok

Det finnes synonymordbøker på norsk, men etter det en kan se, er informasjon om antonymer ganske begrenset.

Valensordbøker

En valensordbok ville gi grunnleggende informasjon om norske verb, og gjennom eksemplermateriale kunne en vise verbene i bruk i enkle og komplekse setninger.

Homografordbok

Det finnes etter det jeg kjenner til, ingen oversikt over antall homografer i norsk, men at antallet ikke er ubetydelig, kan en observere ved å konsultere en vanlig ordbok. I tillegg kommer homografer som oppstår når en tar bøyningsformer med i betraktning.

Fagordlister

Det er behov for å utvikle terminologilister knyttet til ulike områder av yrkeslivet. I dag finnes det en del, men de er delvis mangelfulle med hensyn til hvilken språklig informasjon som gis, med hensyn til volum og i forhold til de ulike praksisfeltene. Slike ordbøker kan gis varierende utforming fra presise definisjonsordbøker til bildeordbøker.

Golden og Hvenekilde (1983) viste i prosjektet "Lærebokspråk" at store deler av vokabularet i grunnskolens fagpensum ikke ble undervist av faglærere med minoritetsspråklige elever. Lærerne ble presentert for fagtekster og indikerte da hvilke ord de anså som "kjente" og hvilke ord som var "fagord", det vil si ord de ville legge vekt på i undervisningen. I fagene fysikk, geografi og historie ble det lokalisert 3432 rotlemma, av disse ble 176 ansett som "kjente", det vil si at de tilhørte det en kan kalle grunnleggende norsk. Av det resterende materialet ville lærerne undervise 1060 fagord. Det gjenstod da 2196 rotlemma som ikke var fagord. Frekvensen av fagord og rotlemma var svært lav. Av fagordene var det bare 74 ord som dukket opp i mer enn ett fag. 50 prosent av ordene utenom de "kjente" dukket bare opp én gang. Disse funnene illustrerer viktigheten av å utvikle vokabularlister med forklaringer til bruk i lærebøker. Store deler av ordmassen i lærebøkene ble altså tatt for gitt av lærerne og ble ikke trukket fram i undervisningen. Det sier seg selv at grunnlaget for å henge med i undervisningen er at en har oversikt over og forstår dette ordmaterialet.

Annet

En kunne tenke seg ulike oversikter over norsk, for eksempel over tekstbindingsmekanismer, tekststrukturer og annet som ville bidra til å utvikle norskkompetanse hos minoritetsspråklige med annen språklig og akademisk bakgrunn. Ved å introdusere slikt beveger en seg bort fra det tradisjonelle ordbokbegrepet, men det betyr ikke at ordsamlinger av denne karakteren ikke er av interesse.

Plattform

Det er selvsagt at nye ordbøker vil foreligge i papirform. Dette bør imidlertid ikke forhindre at ordbøker utgis i andre formater. Ordnett.no tilbyr en rekke ordbøker på nett, og det finnes norsk-norske ordbøker som Bokmålsordboka og Nynorskordboka. En kan også tenke seg ordbøker i form av applikasjoner på nettbrett eller telefon. Det finnes allerede flere slike på

norsk. I takt med at talegjenkjenning er blitt mer robust, er det utviklet ressurser som oversetter direkte og generer oversettelsen til lyd via tekst-til-talesystemer. Presisjonsnivået på oversettelsene er lavt, det vet brukere av Google Translate, men retningen er gitt.

Referanser

- Berkov, V. mfl. 2006. *Ny stor russisk–norsk ordbok*. Moskva : Levende språk
- Berkov, V. 2010. *Ny stor norsk–russisk ordbok*. Moskva : Levende språk
- Golden, Anne, og Hvenekilde, Anne 1983. *Rapport fra prosjektet Lærebokspråk*. Oslo : Universitetet i Oslo, Sentret for språkpedagogikk
- Haugen, Einar 1995. *Norsk–engelsk ordbok*. Oslo: Universitetsforlaget
- Kunnskapsforlaget 2007. *Engelsk blå ordbok*
- Nash, Rose 1997. *Dictionary of Spanish Cognates*. Lincolnwood: NTC Publishing Group
- Rekkedal, Harald mfl. 2005. *Stor norsk–kinesisk ordbok*. Oslo : Kunnskapsforlaget
- Vanvik, Arne 1985: *Norsk uttaleordbok*. Oslo : Fonetisk institutt, Universitetet i Oslo