

STORE TEKSTKORPUS TIL BRUK I SPRÅKRÅDETS SPRÅKOBSERVASJON, SPRÅK- NORMERING OG SPRÅKRÅDGIVNING

ÅSE WETÅS

SAMMENDRAG

Artikkelen tar opp endringene tilgangen til store tekstkorpus har gitt i Språkrådets arbeid med språkoobservasjon og språknormering. Korpusene har gjort det mulig å studere språkbruk og formvalg i store tekstmengder. Faktisk språkbruk (usus) i normdannende tekster skal tillegges vekt i normeringen av begge de to norske skriftspråkene. Tilgang til store tekstkorpus bidrar til å heve kvaliteten og styrke legitimiteten på et arbeidsfelt som har stor oppmerksomhet i offentligheten, og der meningene ofte er både mange og sterke. Artikkelen ser særlig på rollen Leksikografisk bokmålskorpus har fått i den løpende normeringen av bokmålet.

[1] INNLEDNING

Denne artikkelen¹ vil belyse endringene som har skjedd i Språkrådets arbeid med normering og språkoobservasjon som et resultat av at vi etter årtusenskiftet har fått muligheten til å bruke store tekstkorpus i arbeidet vårt. Språkrådet har ansvar for å normere begge de to norske skriftspråkene, men artikkelen legger særlig vekt på normering av bokmålet, og på den viktige rollen ressurser som Leksikografisk bokmålskorpus (LBK) spiller i normeringsarbeidet. I tillegg vil artikkelen også komme nærmere inn på hvordan tilgang til store leksikalske databaser og ressurser bidrar til god kvalitet i arbeidet med andre av Språkrådets kjerneoppgaver. Første del av artikkelen vil være tilbakeskuende og gjøre rede for hvordan både organiseringen av det statlige normeringsarbeidet og den normeringsfaglige praksisen har endret seg fra årtusenskiftet til i dag. Artikkelen vil gjennom eksempler på nyere normvedtak vise hvordan LBK får sin anvendelse i Språkrådets løpende arbeid med å følge opp den norske språkpolitikken på normeringsfeltet.

[1] Artikkelen baserer seg på et foredrag til festsymposium for professor Ruth Vatvedt Fjeld i juni 2018. Den er en takk til Ruth både for det store arbeidet hun har lagt ned i å utvikle Leksikografisk bokmålskorpus, og for hennes mangeårige innsats for norsk leksikografi og leksikologi.

[2] BAKGRUNN

Språkrådet ble etablert i 2005 som statens fagorgan for språk, og rådet er en etat direkte underlagt Kulturdepartementet. Rådet har ingen selvstendige forskningsoppgaver, men forvalter den offentlige norske språkpolitikken gjennom oppgaver som enten er vedtektsfesta, har strategisk prioritet innenfor rammene av den gjeldende språkpolitikken, eller er formulerte i de årlige tildelingsbrevene fra overordnet myndighet. Språkrådet har direkte ansvar for arbeid med norsk, norsk tegnspråk og de nasjonale minoritetsspråkene kvensk, romanes og romani. Rådets oppgaver omfatter både arbeid med språk i offentlig sektor, språk i skole og høyere utdanning, språk i arbeids- og næringslivet og språk på kultur- og mediefeltet. I tillegg har Språkrådet strategisk ansvar for arbeid med norskspråklig fagterminologi og praktisk ansvar for løpende språkbobservasjon og normering av de to skriftspråkene bokmål og nynorsk. Rådet driver et omfattende kurs-, veilednings- og rådgivningsarbeid overfor både offentlige og private aktører på alle våre arbeidsfelt. I tillegg tar vi imot henvendelser fra enkeltpersoner som søker svar på språkspørsmål.

Språkrådets forgjenger var Norsk språkråd (1972-2004). Organisasjonen ble styrt av et faglig råd med 21 representanter for hvert av de to skriftspråkene bokmål og nynorsk, til sammen 42 medlemmer. Rådsmedlemmene representerte til sammen 30 ulike instanser, som omfattet både sentrale språkpolitiske aktører i Norge og relevante fag- og forskningsmiljøer. Norsk språkråd hadde i tillegg ei fagnemnd med fire medlemmer fra hver målform, og et sekretariat av fast ansatte medarbeidere. Norsk språkråd var delt inn i en bokmåls- og en nynorskseksjon, og organisasjonens hovedoppgaver var språknormering og språkrådgivning. Etter hvert fikk rådet også tilsynsoppgaver på mållovsfeltet og strategisk ansvar for utvikling av språkteknologi på norsk. I tillegg hadde rådet i sitt mandat å fremme toleranse og respekt for språk, og å bidra med kunnskap om norsk i offentligheten. I Norsk språkråds oppgaveportefølje lå dessuten ansvar for kvalitetssikring og godkjenning av lærebøker til bruk i skolen. Denne godkjenningsordningen ble avviklet i år 2000.

Språkrådet (2005-) har ansvar for å iverksette den norske språkpolitikken slik den er vedtatt av Stortinget, og på fullmakt fra overordnet departement. Dagens organisasjon har et styre med fire medlemmer oppnevnt av kulturministeren og én representant for de ansatte. Språkrådets styre har overordna økonomi- og resultatansvar for organisasjonen, og styret gjør også løpende normeringsvedtak i tråd med organisasjonens normeringsfullmakter. Språkrådet har i tillegg knyttet til seg fire fagråd der eksperter fra ulike samfunnsfelt er hentet inn for å gi råd og ny kunnskap på organisasjonens arbeidsområder. Ett av fagrådene, fagrådet for normering og språkbobservasjon,

er vedtektsfestet. En hovedoppgave for dette fagrådet er å arbeide med løpende språkobservasjon og språknormering. Jeg vil litt seinere i artikkelen vise fram et par konkrete eksempler på arbeidet til fagrådet for normering og språkobservasjon, for å illustrere hvordan metoder og fagressurser i dette arbeidet har endret seg i den digitale tidsalderen. Men først skal vi ta et nærmere blikk på statusen til og innholdet i dagens bokmålsnorm.

[3] 2005-NORMEN FOR NORSK BOKMÅL

Den gjeldende skriftspråksnormen for norsk bokmål ble vedtatt av Kulturdepartementet i 2004, med iverksettelse fra 2005. Den store, prinsipielle endringen som kom med 2005-reformen var at det gamle tonivåsystemet i bokmålet ble fjernet. Dermed forsvant systemet med hovedformer og sideformer (også kalt «klammeformer»), som hadde kommet inn med 1938-normalen. Systemet med sideformene hadde som sitt primære formål å lage et «venterom» for former på veg ut av normen eller på veg inn i normen, og det fungerte gjennom en periode i norsk normeringshistorie der det langsiktige målet var en gradvis tilnærming mellom de to målformene, med utgangspunkt i folkemålet. For nynorsk ble tonivåsystemet i normen opprettholdt fram til 2012, da dagens nynorskrettskriving ble innført. Opprinnelig var planen på tidlig 2000-tall at begge de to norske målformene skulle gå over til en norm med bare ett nivå i parallelle prosesser, men etter at flertallet i Norsk språkråds fagnemnd stemte mot etnivåsystem for nynorsk, samtidig som det kom tydelige og sterke innvendinger fra høringsinstansene da et forslag til ny nynorsknorm med bare ett normeringsnivå i 2002-2003 var på høring, ble denne endringen utsatt for nynorskens del. Den gamle tonivånormen levde derfor videre, men behovet for en ny og strammere nynorsknorm var tydelig. Språkrådet fikk etter en tid nytt mandat til å sette i verk arbeid med en større rettskrivingsendring for nynorsk. Dette arbeidet ledet fram til det som i dag har fått benevnelsen «2012-normalen for nynorsk».

Det er viktig å presisere at målet med å fjerne tonivåsystemet i bokmålsrettskrivingen i 2005 ikke var å fjerne all valgfrihet innenfor normen. Det var heller ikke et mål i seg selv å reversere resultatene av tilnærminglinja i språknormeringen gjennom 1900-tallet, jf. også Lars Vikør (2007: 202). Vikør skriver at mange tolket opphevingen av tilnærmingssparagrafen i den daværende *lov om Norsk språkråd* som at tilnærmingen mellom målformene skulle reverseres, men han påpeker at det faktisk ikke lå i Stortingets vedtak fra 2002. Selv om en del lite brukte former ble tatt ut av normen, er det fremdeles en god del variantformer i bokmålet, men fra 2005 har disse formvariantene vært formelt sidestilte, slik at de i prinsippet skal ha samme status i bruk. I praksis ser vi likevel at systemene for implementering av hele innholdet i de offisielle språknormene ikke

er gode nok, og mange store private normaktører (forlags- og mediehus) viser i liten grad fram den faktiske variasjonsbredden i bokmålet. Mange brukere kjenner derfor dessverre ikke til valgmulighetene de har innenfor normen, og en del valgfrie former er derfor mindre synlige i offentligheten enn det som er ønskelig.

Bokmålsnormen fra 2005 bygger på normeringsvedtak gjort av Norsk språkråd i perioden 2002-2004. Som nevnt forsvant en del lite brukte former ut av bokmålet på dette tidspunktet, samtidig som det ble tatt inn en del nye brukte former som enten hadde vært sideformer, eller som hadde ligget utenfor normen. I 2002 hadde § 1 i *lov om Norsk språkråd* blitt endret av Stortinget, slik at loven ikke lenger la til grunn at det langsiktige målet for språknormeringen var å fremme utviklingstendenser som førte bokmål og nynorsk nærmere hverandre. Fra tidlig 2000-tall blir de to norske målformene dermed normerte på selvstendig grunnlag og med utgangspunkt i egen tradisjon. Dagens retningslinjer for normering av bokmål og nynorsk presiserer samtidig at det i normeringen av de to målformene skal unngås unødige forskjeller i ortografiske prinsipper (f.eks. i stavemåten av lånord).²

Hvordan skaffet språknormererne på 1900-tallet seg innsikt i hvordan brukspråket så ut og hvilke former som var mye og lite brukte, og hvilke kilder brukte de? I arbeidet med 2005-normen for bokmål ble det i stor grad fremdeles brukt tradisjonelle metoder for å hente inn empiri om faktisk språkbruk. Et eksempel er de to større studiene Norsk språkråd på 1990-tallet bestilte av normuttrykket i norsk skjønnlitteratur (bokmål og nynorsk). I disse studiene ble 59 romantekster fra ulike tiår på 1900-tallet gjort maskinlesbare – noe som for så vidt representerte et visst framskritt i det empiriske grunnlagsarbeidet, sammenliknet med tidligere større rettskrivingsendringer. Maskinlesingen skulle gjøre det mulig å kartlegge formvariasjon i en større mengde tekster under ett. Lars Vikør (1995) skriver at de to parallellstudiene ble initierte som en konsekvens av motstand mot 1981-reformen for bokmål. Motstanden hadde utgangspunkt i at det ikke var gjort større empiriske undersøkelser av den faktiske utviklingen av bokmålet gjennom 1900-tallet. Det var Språkrådet som stod for utvalget av de 59 romantekstene, og for bokmål kom Egil Pettersens studie «Språknormering og forfatterne» i 1993. I denne studien undersøkte Pettersen ortografi og morfembruk i skjønnlitterære tekster fra ti utvalgte bokmålsforfattere fra hvert av årene 1937, 1957 og 1977. Lars Vikør (1995): «Rettskriving hos nynorskforfattarar» er en tilsvarende undersøkelse og analyse av 29 nynorskforfatteres tekster fra de samme tre tiårene.

Egil Pettersen skriver i forordet til sin studie at «Ideelt sett burde alle typer

[2] Jf. Språkrådets retningslinjer for normering, <https://www.sprakradet.no/globalassets/sprakavare/norsk/retningslinjer-for-normering-bokmalsversjon.pdf>

litteratur vært med i en undersøkelse som skulle være fullt ut representativ for språkbruken. Vi hadde gjerne sett at aviser, ukeblad og tidsskrifter også hadde fått plass i prosjektet. Men det ville ha sprengt alle rammer og ligget utenfor våre muligheter» (Pettersen 1993: 1). I sluttordet til studien peker Pettersen (1993: 163) på at materialet han har analysert, er spinkelt og vanskelig å trekke slutninger fra. Dette har selvsagt sammenheng med at et utvalg på til sammen 30 tekster er altfor lite til at man kan finne sikre og gode funn på alle undersøkelsesområder, samtidig som det blir svært synlig at materialet mangler nødvendig sjangerspredning. Pettersens og Vikørs studier er solide og grundige, men metoden med i hovedsak manuell kartlegging av ortografi og morfembruk er ekstremt arbeidskrevende. Det er samtidig viktig å understreke at de to studiene ikke hadde som intensjon å kartlegge særlig fremragende bruk av språket, men å undersøke hvilket gjennomslag rettskrivingsreformene i 1917, 1938 og 1959 hadde hatt i mye leste tekster. Med andre ord var forutsetningen for studiene ikke å bygge den videre språknormeringen direkte på de utvalgte forfatternes formvalg, men snarere å dokumentere om og hvordan språkbruken i de skjønnlitterære verkene hadde blitt påvirket av innholdet i de respektive rettskrivingsreformene.

Samme modell og undersøkelsesmetoder som hos Pettersen (1993) og Vikør (1995) lå også til grunn for to av i alt fire studier jeg selv leverte Språkrådet i 2001 (Wetås 2001a, b). Disse studiene kartla formvalg i ulike typer publisert og upublisert nynorsktekst, og de skulle utgjøre underlagsmateriale til det da pågående arbeidet med ny nynorsknorm. Selv om det på dette tidspunktet faktisk var utviklet noen relativt små digitale tekstkorpus for norsk, foregikk kartleggingsarbeidet stadig hovedsakelig i form av mer eller mindre manuell ekserpering fra lokalaviser, NOU-er og utvalgte sakprosaverk.

Her kan det skytes inn at det selvsagt også ville vært prinsipielt mulig å bygge språknormeringen på skriftspråksuttrykket til et utvalg fremragende forfattere, eller på skjønnet til noen utvalgte spesialister, men det ville representere en framgangsmåte og en normeringspolitikk som står i motsetning til det som har ligget til grunn for den norske språknormeringstradisjonen. Gjennom hele 1900-tallet var tilnærminga mellom bokmål og nynorsk på folkemålsgrunn det langsiktige formålet med språknormeringen. Dermed ligger det i sakens natur at mye brukte former i det norske folkemålet (talt norsk) skulle tillegges særlig vekt i normeringssammenheng, og at tonivåsystemet hadde en viktig funksjon i å få realisert dette målet.

Selv om tilnærminglinja mellom de to målformene i praksis alt var forlatt, førte Stortingets enstemmige vedtak fra 2002 om at bokmål og nynorsk skulle

normeres på egen grunn, og uten sideblikk til den andre målformen, til en endring i vektleggingen av normeringsprinsippene som har vært brukt for de norske skriftspråkene. En annen viktig endring i betingelsene for den skriftspråklige normeringen av norsk på 2000-tallet, er den radikale økningen i tilgang på store mengder data om faktisk språkbruk i ulike typer brukstekster. Den sistnevnte utviklingen kom raskt til å endre kravene til empirisk fundament for normeringsvedtak. I tillegg kunne åpne dataløsninger ha den gode tilleggs-effekten at det sikret allmenn tilgang til å kikke språknormererne i korta, og dermed gav åpne, tilgjengelige korpus språkbrukerne mulighet til å vurdere i hvilken grad retningslinjene for normering ble etterlevd av normererne. Når dette er sagt, er usus (språkbruk) bare ett av til sammen sju prinsipper for normering av bokmål (jf. nedenfor). Retningslinjene legger dermed ikke opp til at det skal være et entil-en-forhold mellom bruksfrekvens i normdannende tekster og faktisk representasjon i skriftspråksnormen.

[4] SPRÅKRÅDETS NORMERINGSFULLMAKTER OG NORMERINGSPRINSIPPER

Etter at Språkrådet var etablert, gav Kulturdepartementet i 2012 organisasjonen nye normeringsfullmakter. Fra dette tidspunktet har det offentlige normeringsoppdraget vært nedfelt i Språkrådets vedtekter. Paragraf 3 *Normering av bokmål og nynorsk* slår i første ledd fast at Språkrådet kan gi råd om staving og bøyning av tidligere unormerte ord og gjøre bindende normeringsvedtak når språkutviklinga tilsier det. I tredje ledd slås det fast at Språkrådet også kan vedta ny stavemåte og bøyning av tidligere normerte ord, dersom faktisk språkbruk eller systemhensyn tilsier det. Det ligger likevel klare avgrensninger i Språkrådets normeringsfullmakter. De gjelder nemlig bare for normering av enkeltord eller grupper av ord under systemnivå. Alle *gjennomgripende* endringer eller *systemendringer* i de to norske skriftspråksnormene er det Kulturdepartementet som skal godkjenne, etter forslag fra Språkrådets styre.

Språkrådets fagråd for normering og språkbobservasjon leverte i 2014 utkast til interne retningslinjer for normering av både bokmål og nynorsk, som styret i Språkrådet vedtok med virkning fra 1. mars 2015. Normeringen av bokmål og nynorsk følger etter disse retningslinjene samme hovedprinsipper, men med en litt ulik vektlegging for de to målformene. I normeringen av bokmålet skal følgende prinsipper vektlegges:

- (i) **Selvstendighetsprinsippet:** Målformen skal normeres på egen tradisjon og eget grunnlag, men ikke slik at det oppstår nye og unødvendige forskjeller mellom de to målformene når det gjelder ortografi, valg av norvagserte

former osv. Det er f.eks. ingen grunn til at ord som *kviss* og *nikab* skal skrives ulikt på de to målformene.

- (ii) **Stabilitetsprinsippet:** Normen skal ligge fast over tid, og det kan bare gjøres mindre justeringer. Dette skal gi språkbrukerne tilgang til en langsiktig og stabil norm som skal være enkel å gjøre seg kjent med og lett å kunne bruke over tid.
- (iii) **Bruksprinsippet (ususprinsippet):** For bokmål er det vesentlige grunnlaget nettopp skriftspråket slik det kommer til uttrykk i tekster skrevet på bokmål, og da særlig tekster som regnes som normdannende. Vektleggingen av «normdannende tekster» kan problematiseres, for en rigid bruk kan gjøre normeringsprosessen sirkulær. Språkrådets fagråd for normering og språkbetjening startet i 2019 derfor et arbeid med å presisere normeringsprinsippene på dette punktet.
- (iv) **Stramhetsprinsippet:** Det skal ikke være et selvstendig mål med normeringen å stramme inn eller utvide valgfriheten i bokmålsnormen. På lang sikt sier retningslinjene at det likevel bør vurderes om normen kan strammes inn *noe*. Her gjelder det likevel at Språkrådet bare har en avgrensa normeringsfullmakt og ikke på eget initiativ kan gjøre mer gjennomgripende endringer.
- (v) **Enkelhetsprinsippet:** Det bør være færrest mulig bøyingsmønstre, og normeringen skal heller gjelde klasser eller grupper av ord, snarere enn enkeltord. Det er ikke vanskelig å se at dette prinsippet kan støte mot bruksprinsippet, og her legger retningslinjene faktisk inn en avgrensing i favør av bruksfrekvens. Hvis høyfrekvente ord har et avvikende bøyingsmønster, bør dette mønsteret stå ved lag.
- (vi) **Tradisjonsprinsippet:** Både bokmål og nynorsk har mer enn hundre års skrifttradisjon bak seg, og det gjør at mye leste eldre tekster også bør være med i det normdannende tekstgrunnlaget for skriftspråket. Dette prinsippet knytter seg tett til stabilitetsprinsippet.

For nynorsk gjelder i tillegg **talemålsprinsippet**. Det slår fast at ved innstramning av normen skal formene som har best grunnlag i talemålet stå igjen, mens det ved eventuelle utvidinger av normen helst skal velges former som har et solid talemålsgrunnlag.

For at normeringsprinsippene skal kunne praktiseres på en god måte, er det

nødvendig for Språkrådet å ha tilgang til store mengder relevant empiri som kan gi informasjon om hvordan språket i bruk faktisk ser ut, og hvordan det har endret seg over tid. Av de nevnte normeringsprinsippene for norsk bokmål forutsetter i alle fall både bruksprinsippet, enkelhetsprinsippet og tradisjonsprinsippet tilgang til kvalitativt og kvantitativt solid empiri.

Min egen bakgrunn fra leksikografisk arbeid og som medlem i nemnda som i 2011 leverte forslag til ny nynorskrettskriving, har vist meg veldig klart at man i normeringssammenheng vanskelig kan stole på introspeksjon og egen intuisjon når det gjelder hvilke formvarianter som er i faktisk bruk, og hvilke varianter som dominerer i bredt sammensatte tekstmengder. Det er bare store tekstsamlinger som vil kunne gi sikker innsikt på dette punktet, og da vil Leksikografisk bokmålskorpus (LBK) og andre store digitale tekstsamlinger være helt nødvendige arbeidsverktøy i den løpende normeringen av de to norske skriftspråkene.

[5] BRUKEN AV LEKSIKOGRAFISK BOKMÅLSKORPUS (LBK)³ OG ANDRE STORE DIGITALE KORPUS

Språkrådets styre står for de endelige vedtakene i normering av enkeltord i bokmål og nynorsk. Språkrådets fagråd for normering og språkbobservasjon leverer saksunderlag og anbefalinger til ny eller endret normering til styret, og saksunderlagene gjør grundig greie for det empiriske grunnlaget for fagrådets anbefalinger. Fagrådet tar løpende opp normeringsspørsmål som trenger avklaring, men de arbeider i tillegg også tematisk. De tre siste årene har Språkrådets fagråd for normering og språkbobservasjon arbeidet med å kartlegge behovet for nedskjæring i genusvariasjonen innafor begge de to skriftnormene. På oppdrag fra Språkrådet gjorde forsker Philipp Conzett ved Universitetet i Tromsø i 2016 en empirisk studie av genusvariasjon i norsk (Conzett 2017). I dette arbeidet var LBK og Det nynorske tekstkorpuset, begge utviklet for leksikografiske formål ved Universitetet i Oslo, viktige kilder til informasjon om bruken av de ulike genusvariantene. I 2018 satte fagrådet dessuten i gang et prosjekt for å undersøke nærmere statusen til norvagiseringsvedtak eldre enn 25 år. Også i dette arbeidet vil det være helt avgjørende å ha tilgang til store moderne tekstsamlinger som kan vise i hvilken grad norvagiserte skrivemåter er tatt i bruk.

For å vise den konkrete bruken av LBK og andre sentrale og nyere datakilder i normeringsarbeidet vårt, vil jeg vise fram underlagsmateriale fra to normeringssaker som ble behandlet av Språkrådets styre i 2016 og 2017, og der styret

[3] Leksikografisk bokmålskorpus (LBK) er et balansert leksikografisk korpus på ca. 100 mill. ord, og det inneholder tekst fra perioden 1985-2013. Sammensetningen av korpuset skal være representativ for både ulike typer tekster og mengden av disse teksttypene en gjennomsnittlig leser møter. I presentasjonen av korpuset (<https://www.hf.uio.no/iln/om/organisasjon/tekstlab/prosjekter/lbk/>) går det fram at 95 % av tekstene som inngår, er redaksjonelt bearbeida og publiserte av forlagshus og mediehus.

gjorde vedtak om endret rettskriving. Det må presiseres at eksemplene omfatter bare to del saker i det som er større sakskomplekser, og de utgjør slik sett en liten del av det samla normeringsarbeidet til rådet.

[5.1] *Normeringen av substantivene mormon og mormoner*

Både i bokmål og nynorsk var formen *mormon* eneform fram til 2017, men vi ønsket å vurdere om normen burde justeres. Langformen *mormoner* var rett nok inne i bokmålsrettskrivingen fra 1984 til 1986, men ble tatt ut igjen, som en konsekvens av at Jesu Kristi kirke av de siste dagers hellige *selv* bare brukte kortformen. Da fagrådet for normering og språkoobservasjon i 2017 behandlet saken, mente de at det ikke kunne legges avgjørende vekt på hva trossamfunnet selv foretrekker. Ønsker fra enkeltaktører ligger utenfor settet av normeringsprinsipper som gjelder for norsk, og kan ikke uten videre tas hensyn til.

Søk i korpus gir interessante funn om bruken av kortformen *mormon* (KF) og langformene *mormoner* og *mormonar* (LF), jf. tabell 1 (nedenfor).

		KF	LF	n ⁴
nb.no (bøker)	Nasjonalbibliotekets korpus av digitaliserte bøker og aviser, 1960–	47 %	53 %	1 357
nb.no (aviser o.a.)	Norske aviser 1998–2015 (bokmålsdelen)	28 %	72 %	1 917
Norsk aviskorpus	Balansert korpus av bokmålstekster, 1985–2013	25 %	75 %	267
Leksikografisk bokmålskorpus	Nynorsktekster fra ulike sjangre fra 1800-tallet til i dag	67 %	33 %	48
Nynorskkorpuset	Norske nettsider 2009–2010	76 %	24 %	17
NoWaC		30 %	70 %	139

TABELL 1: Oversikt over korpus treff på søk «mormon», «mormonar» og «mormoner»

Her kommer en tydelig forskjell til uttrykk mellom LBK og nynorskkorpuset på den ene sida og de andre korpusene på den andre sida. Begge de to leksikografiske korpusene har en sterk overvekt av kortformen *mormon*. En forklaring på at disse to korpusene skiller seg ut, kan være at en større andel av teksttilfanget der er normert etter forlags- og mediehusnormer enn tilfanget i de andre korpusene (og da særlig NoWaC). Med den bakgrunnen kan det gi ekstra styrke til de totale tallene at både LBK og Nynorskkorpuset likevel omfatter en relativt stor mengde forekomster av en formvariant som ligger utenfor den offisielle skriftspråksnormen. Denne entydige empiriske dokumentasjonen var en del av bakteppet for det seinere vedtaket om å ta opp skrivemåten *mormoner* ved siden

[4] n = totalt antall korpus treff. Tabellen er hentet fra saksframlegget til styresak 32/17, behandlet av Språkrådets styre 2.6.17.

av *mormon* som en del av den offisielle normen.

[5.2] Normeringen av ordene *dølahest*, *dølafe* og *dølaku*

I bokmålsnormen var *dølehest*, *dølefe* og *døleku* (med *e*-fuge) eneformer fram til 2016. Språkrådets fagråd for normering og språkoobservasjon så nærmere på denne ordgruppa, og søk i korpus gav også her interessante funn.

		<i>dølahest</i>	<i>dølehest</i>
nb.no (bøker)	Nasjonalbibliotekets korpus av digi-	55 %	45 %
nb.no (aviser o.a.)	taliserter bøker og aviser, 1960–	60 %	40 %
Retriever	Norske aviser og andre media	60 %	40 %
Norsk aviskorpus	Norske aviser 1998–2015	48 %	52 %
Leksikografisk	Balansert korpus av bokmålstekster	56 %	44 %
bokmålskorpus	fra 1985 til i dag		
NoWaC	Norske nettsider 2009–2010	42 %	58 %
Google	Google.no	83 %	17 %

TABELL 2: Oversikt over korpustreff på søk «*dølahest*» og «*dølehest*»⁵

Også for disse ordene, i tabell 2 representert ved *dølahest* og *dølehest*, viste det seg at formvarianten som ikke var en del av den offisielle skriftspråksnormen, var svært mye brukt. Overraskende nok hadde den ikke-normerte varianten med *a*-fuge faktisk flere treff både i LBK og i Nasjonalbibliotekets digitale bokhylle enn den normerte formen med *e*-fuge.

I normeringssaker er det også andre forhold enn bruksfrekvens som skal vurderes, ikke minst hensynet til stabilitet og system (jf. oversikten over normeringsprinsipper). Det endelige vedtaket i saken ble derfor ikke å fjerne de tradisjonelle formvariantene med *e*-fuge (som faktisk også viste seg å ha en god del bruk bak seg) til fordel for de mer brukte med *a*-fuge. Løsningen ble i stedet å ta de sistnevnte formene (*dølahest*, *dølafe*, *dølaku*) inn i normen, og la dem stå som likestilte med *dølehest*, *dølefe* og *døleku*.

[6] BRUK AV LBK TIL SPRÅKOBSERVASJON OG NORMERING

Sammenlignet med andre store tekstkorpus er en umiddelbar fordel med Leksikografisk bokmålskorpus at det er representativt og vektet, og at det omfatter tekst fra en hel rekke ulike sjangre. Jeg vil i denne sammenhengen særlig løfte fram sjangeren TV-teksting. Denne teksttypen representerer en vesentlig kilde til kunnskap om skriftlig norsk for mange som i liten grad leser andre typer trykt eller digital tekst. I språknormering er derfor dette en viktig

[5] Tabellen er hentet fra saksframlegg til styresak 40/16, behandlet av Språkrådets styre 23.8.16.

teksttype. LBK dekker tidsperioden fra 1985 til 2013, og vektingen i korpuset gir i det hele tatt et godt bilde av hvilke normframstillinger språkbrukerne møter i det daglige. Tekstkorpuset viser selvsagt ikke hvilke formvarianter språkbrukere flest foretrekker når de skal produsere egne tekster, men det gir god og kvalitetssikret dokumentasjon på hvilke operative normer forlag, mediehus og andre store tekstprodusenter bruker i sine tekster, og dermed hvilke formvarianter språkbrukerne faktisk eksponeres for. LBK dokumenterer at dette i stor grad viser seg å være former som ligger i det som ofte omtales som den «konservative» eller «moderate» enden av bokmålsnormen, slik også Kjersti W. Kola har dokumentert i masteroppgaven sin fra 2014. Det er selvsagt grunn til å tro at disse normframstillingene i tur preger språkbrukernes oppfatninger av hva som er rammene for det korrekte, normerte språket. Svaret er likevel at vi faktisk vet veldig lite om hvilke former språkbrukerne foretrekker eller oppfatter som del av normen, annet enn at det er åpenbart at språkbrukere velger ulike normuttrykk i ulike brukssituasjoner. For eksempel viser en befolkningsundersøkelse Opinion utførte for Språkrådet i 2016 at svært mange voksne språkbrukere rapporterer å bruke dialekt når de skriver i sosiale medier. Denne tendensen er særlig uttrykt i responsen fra språkbrukere i Nord-Norge og Midt-Norge, den er tydeligere for kvinner enn for menn, og den gjelder yngre språkbrukere i større grad enn eldre. Det skal skytes inn at undersøkelsen er basert på egenvurdering av skriftspråklig praksis, så hva den enkelte respondenten legger i det å skrive på dialekt, kan være ulikt. Vi ønsker oss mer kunnskap på dette området, men studier av unormert og upublisert tekst produsert av «vanlige» språkbrukere (i motsetning til profesjonelle språkbrukere) er vanskelig å gjennomføre. En hovedgrunn er at rettighetsproblematikk og personvernlovgivning hindrer innhøsting og bruk av tekst fra sosiale medier. En forskningsbasert undersøkelse av språkbrukernes praktiske normkunnskaper og (sjangerbestemt) normvariasjon står likevel høyt på Språkrådets ønskeliste.

For Språkrådet er LBK av stor nytteverdi i det løpende arbeidet med språkbobservasjon og normering, og det er også verdifullt i arbeidet med språkrådgivning. I tillegg brukes LBK i arbeidet med vedlikehold og utvikling av *Bokmålsordboka* og *Nynorskordboka*, som Språkrådet fra 2016 eier sammen med Universitetet i Bergen (UiB). I et prosjekt med oppstart i 2018 og tentativ slutføring i 2023 skal de to tvillingordbøkene gjennomgå en fullstendig revisjon. Prosjektet er finansiert av Kulturdepartementet, med bidrag fra UiB og Språkrådet. Viktige delprosjekter i dette arbeidet vil være kritisk gjennomgang og revisjon av alle orddefinisjoner og brukseksempler. Et korpus spesielt utviklet til bruk i leksikografisk

forskning er av stor verdi i dette arbeidet, og det utgjør en sentral ressurs i revisjonsprosjektet. Vi vet at språket på noen områder endrer seg raskt, og til arbeid med Språkrådets rettskrivningsordbøker skulle vi gjerne sett at de viktige leksikografiske korpusressursene for norsk løpende ble bygget ut med nytt teksttilfang.

[7] OPPSUMMERENDE MERKNADER

Språkrådet er et forvaltningsorgan uten forskningsakkreditering. For å kunne utføre arbeidsoppgavene våre, må vi ha inngående kjennskap til språk og språkendring. I dette arbeidet er vi avhengige av tett kontakt med relevante forskningsmiljøer innenfor fagfelt som leksikografi og leksikologi, onomastikk og terminologi. Det er mye vi ønsker mer kunnskap om. Er det for eksempel endring i den skriftspråklige bruken av *han* og *ham* i bokmål? Opptrer objektsformen *dykk* stadig oftere på subjekt plass i tekster på nynorsk, og hvordan kan det i så fall forklares? Har utbredte trekk fra norsk dialekt syntaks som «*Hvem bil?*» gjennomslag også i skrift? Hva er utsiktene for at et nytt kjønnsnøytralt personlig pronomen *hen* vil etablere seg i skriftspråket? Dette er bare et lite utvalg eksempler på spørsmål der det er viktig for Språkrådet å ha tilgang til relevant, oppdatert empiri. Uten tilgang til store og kvalitetssikra korpusressurser vil det være tilnærmet umulig å få oversikten vi trenger over hvordan språket i bruk faktisk tar seg ut, både når det gjelder ortografi, morfologi og syntaks.

Alle Språkrådets arbeidsfelt er avhengige av god tilgang til forskning og forskningsinfrastruktur utviklet og forsket fram av relevante forskningsmiljøer. Vi trenger dessuten en tett og god dialog med forskningsmiljøene for å utføre forvaltningsoppgavene våre på beste vis. Språkrådet ser derfor med bekymring på at den systematiske forskningen i det norske språkets leksikon lever en pressa og utrygg tilværelse ute i universitets- og høgskolesektoren. For forskningsfeltene leksikografi og leksikologi vurderer vi at situasjonen nå er prekær. Det samme gjelder for de leksikalske «nabodisiplinene» terminologi og stedsnavnforskning. Vi ser dessuten at det ikke bare er forskningsfeltene i seg selv som er under press, men at det også er svært krevende å skaffe nødvendige ressurser til å opprettholde og videreutvikle relevant forskningsinfrastruktur på de nevnte fagområdene. Disiplinene deler skjebne med en rekke andre såkalte nasjonale empirier på kulturfeltet, og det er nærliggende å se utviklingen i sammenheng med finansieringsmodellene i sektoren. Språkrådet engasjerer seg i disse spørsmålene for å sikre at det norske samfunnet og alle andre som ønsker det, skal ha tilgang til relevant infrastruktur og forskningsressurser for norsk språk. For at vi skal kunne dokumentere hvilken relevans dette har for oss som fagorgan, forsøker vi også selv å løfte fram og synliggjøre den konkrete

nytteverdien Språkrådet har av både fagkompetansen og de enkelte leksikalske infrastrukturressursene fra forskningssektoren. I dette bildet er Leksikografisk bokmålskorpus én av flere tunge og viktige ressurser, og korpuset bidrar både til å sikre god datakvalitet og gjøre Språkrådets arbeid med språkobservasjon og språkrådgivning mer effektivt.

LITTERATUR

- Conzett, Philipp. 2017. *Genusvariasjon i norsk skriftspråk*. Ei undersøking av genusbruken ved substantiv som er normerte som maskulinum og nøytrum. (Rapport utarbeidet for Språkrådet.) <https://septentrio.uit.no/index.php/SapReps/article/view/4077>.
- Kola, Kjersti W. 2014. *Bokmålsbruk – hvorledes/hvordan/åssen og hvorfor? Om bruken av morfologiske og ortografiske varianter i bokmålsnormalen*. Masteroppgave i nordisk, særlig norsk, språkvitenskap, Universitetet i Oslo. <https://www.duo.uio.no/bitstream/handle/10852/40996/Kola-Master.pdf?sequence=1&isAllowed=y>.
- Pettersen, Egil. 1993. *Språknormering og forfatterne. Ortografi og morfembruk hos ti bokmålsforfattere for hvert av årene 1937, 1957 og 1977*. Eigenproduksjon nr. 49/93. Bergen: Nordisk institutt, Universitetet i Bergen.
- Språkrådets retningslinjer for normering av bokmål og nynorsk*. <https://www.sprakradet.no/globalassets/spraka-vare/norsk/retningslinjer-for-normering-bokmalsversjon.pdf>
- Vikør, Lars. 1995. *Rettskriving hos nynorskforfattarar*. Norsk språkråds skrifter nr. 3. Oslo: Norsk språkråd.
- Vikør, Lars. 2007. *Språkplanlegging*. Prinsipp og praksis. Oslo: Novus forlag.
- Wetås, Åse. 2001a. Undersøking av formbruken i ulike typer upublisert materiale på nynorsk. (Rapport utarbeidet for Norsk språkråd.) <https://www.sprakradet.no/localfiles/nyno02awe3.pdf>
- Wetås, Åse. 2001b. Undersøking av formbruken i ulike typer publisert materiale på nynorsk. (Rapport utarbeidet for Norsk språkråd.) <https://www.sprakradet.no/localfiles/nyno02awe2.pdf>

KONTAKT

Åse Wetås

direktør, Språkrådet

ase.wetas@sprakradet.no