

Tilhøvet mellom fastsette og operative normer i bokmål

Av Stig J. Helset

Samandrag

Denne artikkelen undersøker korleis valfridomen innanfor offisielt bokmål nedfeller seg i operative normer i eit stort og breitt samansett utval av redigerte bokmålstekstar. Hausten 2021 vedtok styret i Språkrådet nye retningslinjer for normering av bokmål og nynorsk. I desse retningslinjene heiter det at «[d]et vesentlege grunnlaget for normeringa av bokmål er slik det ser ut i bokmålstekstar», og at normeringsgrunnlaget i første rekkje skal vere «tekstar som har vore gjennom ein redigeringsprosess». Med det til siktemål å styrke kunnskapsgrunnlaget for den vidare normeringa av bokmål har vi derfor gjennomført ein studie der vi har kartlagt bruken av ulike valfrie ordformer mellom 150 språklege variablar i eit breitt samansett utval av 452 redigerte og publiserte tekstar som ligg i Leksikografisk bokmålskorpus. I tillegg til registrering og komparasjon av bruksfrekvensar for ulike valfrie former innanfor einskildvariablar, har vi brukt statistisk korrespondanseanalyse for å kartlegge normklynger av ulike subvarietetar av bokmål. Hovudkonklusjonen er at dei moderate formene dominerer over dei radikale formene for dei aller fleste einskildvariablane som er med i studien, samstundes som dei radikale formene dominerer over dei moderate for nokre få av einskildvariablane, og at det mellom tekstane i utvalet som er med i studien, eksisterer ei operativ normklynge som ein kan karakterisere som moderat bokmål.

1 Innleiing¹

Spørsmålet om kva som bør vere normgrunnlaget for offisielt skriftspråk i Noreg, har vore omstridd heilt sidan unionsoppløysinga med Danmark i 1814. Etter vedvarande og tidvis heftige stridar om normgrunnlaget for og ikkje minst normeringa av dei to offisielle norske skriftspråka vedtok Stortinget ved inngangen av 2000-talet at tilnæringslinja ikkje lenger skulle vere gjeldande politikk. Tvert imot slo stortingsmeldinga om målbruk i offentleg teneste (Kulturdepartementet 2001–2002: 25) fast at måldyrkingsarbeidet på 2000-talet skal vere «knytt til spørsmålet om korleis bokmål og nynorsk kvar for seg skal kunne utvikle seg best mogleg», og på dette grunnlaget blei det vedteke ny rettskriving for bokmål i 2005 og for nynorsk i 2012. I denne artikkelen skal vi rette søkjelyset mot bokmålsnormalen og praktiseringa av denne i bokmålstekstar.

Jamført med rettskrivingane av 1917, 1938 og 1959 kan gjeldande rettskriving for bokmål seiast å vere kjenneteikna av fire hovudtrekk. For det første er 2005-normalen altså danna på sjølvstendig grunnlag, uavhengig av nynorsk. For det andre er det tidlegare skiljet mellom hovudformer og sideformer oppheva, slik at ein no berre har anten eineform eller to eller fleire jamstilte former ved staving og bøying av eit ord. For det tredje har ein tatt ut ei rekkje radikale former som var tillatne, og jamvel gjerne hovudformer, i tidlegare rettskrivingar. For det fjerde har ein opna opp for ei lang rekkje tradisjonelle og konservative skriftformer som anten ikkje var tillatne i det heile eller berre tillatne som sideformer i tidlegare rettskrivingar. Det siste trekket representerer elles ei vidareføring av den liberaliseringslinja som blei trekt opp i og med bokmålsrettskrivinga av 1981 (Guttu 2005: 2).

Hausten 2021 vedtok styret i Språkrådet (2021) nye *Retningslinjer for normering av bokmål og nynorsk*, og i desse retningslinjene heiter det at «[d]et vesentlege grunnlaget for normeringa av bokmål er språket slik det ser ut i bokmålstekstar», og at normeringsgrunnlaget i første rekke skal vere «tekstar som har vore gjennom ein redigeringsprosess»

1. Takk til Helge Dyvik, som har vore ein viktig medspelar når det gjeld tekstutval og utval av språklege variablar i denne studien, og til Paul Meurer, som har stått for det tekniske uttrekket av tekstar frå Leksikografisk bokmålskorpus og for frekvensregistrering av språklege variantar i korpuset. Takk også til to anonyme konsulentar og redaksjonen i *Maal og Minne* for konstruktive tilbakemeldingar som har bidrege til at dette har blitt ein betre artikkel.

(Språkrådet 2021: 12). I denne artikkelen vil vi *ikkje* drøfte i kva grad desse retningslinjene representerer ein føremålstenleg språkpolitikk. Vi berre konstaterer at ifølgje retningslinjene *skal* eventuelle justeringar av bokmålsnormalen i første rekkje byggje på dette bruksprinsippet (Språkrådet 2021: 12), og tek dette som utgangspunkt for artikkelen. Retningslinjene slår vidare fast at «[v]urderingar på bakgrunn av bruksprinsippet må byggja på undersøkingar av store korpus av bokmålstekstar som har god sjangerspreiing og regional spreieing, og som både speglar dagens språk og går nokre tiår tilbake i tid.» Dessutan kan ein lese i retningslinjene at «[e]it resultat av tilnæringsperioden er at skriftleg bokmål i dag viser større variasjon mellom alternative ordformer enn det som elles er vanleg i moderne skriftspråk», og at «[t]ekstmaterialet viser ein del samvariasjon mellom dei ulike formvala, slik at vi kan snakka om operative subnormer av bokmål i form av varietetar som tradisjonelt har vore kalla «konservativt», «moderat» og «radikalt» bokmål, med mellomformer» (Språkrådet 2021: 11). Det blir vidare vist til proposisjonen for Språklova som tredde i kraft 1.1.2022 (Kulturdepartementet 2019– 2020: 22):

Rettskrivingsreforma for nynorsk i 2012 var tufta på ein grundig gjennomgang av bruksmønster for mange grupper ord. Undersøkingane baserte seg på store tekstkorpus som gjev eit godt bilete av bruken i moderne nynorsk. Den siste store rettskrivingsreforma for bokmål kom i 2005, og då var det ikkje høve til å gjere tilsvarende undersøkingar av bruken i bokmålstekstar. Det kan derfor vere grunn til å sjå på om det trengst justeringar i bokmålsnorma i lys av bruken i dag slik han framstår i store korpus med bokmålstekst. Det vil likevel ikkje vere aktuelt å gjere gjennomgripande endringar i bokmålsnorma.

På bakgrunn av denne offisielle normeringspolitikken har vi gjennomført ein studie der vi har kartlagt bruken av ulike valfrie ordformer mellom 150 språklege variablar i eit breitt samansett utval av 452 redigerte og publiserte tekstar som ligg i Leksikografisk bokmålskorpus (LBK), for slik å styrke kunnskapsgrunlaget for eventuelle justeringar av bokmålsnormalen. Det overordna forskingsspørsmålet for studien var om det i tekstutvalet som ligg til grunn for studien, eksisterer *normklynger* beståande av bestemte ordformer som kunne la seg identifisere som meir eller mindre dominerande subvarietetar av bokmål. Vi definerte ei «normklynge» som «eit sett av språklege formval som er felles for eit antal tekstar på ein slik måte at tekstane skil seg som gruppe frå andre

tekstar» (jf. Helset 2017: 12). Med bakgrunn i funn frå ein studie av Dyvik (2012: 218) som indikerer klyngestrukturar i eit utval avistekstar, sette vi fram ein hypotese om at det innanfor tekstutvalet frå LBK ville vekse fram normklynger av bestemte ordformer som kunne la seg identifisere som statistisk signifikante subvarietetar av bokmål.

2 Teoretisk bakgrunn

2.1 Fastsette og operative normer

Dei *fastsette* skriftspråksnormene for nynorsk og bokmål blir forvalta og fastsette av Språkrådet, medan gjennomgripande endringar skal godkjennast av Kulturdepartementet (Kulturdepartementet 2019–2020: 22). Sjølv om dei fastsette normene i prinsippet omfattar alle dei grammatiske nivåa, er det i første rekkje det som gjeld staving og bøying, Språkrådet gjer eksplisitte vedtak for (Vedtekter for Språkrådet, §3), og det er nettopp den ortografiske og morfologiske valfridomen i bokmål som står i søkjelyset i denne artikkelen. Vi skal nemleg sjå nærmare på i kva for grad denne valfridomen har gjeve seg utslag i *operative* normer i eit stort og breitt samansett utval av redigerte bokmålstekstar i LBK. Dyvik (2009: 111) definerer operative normer som «en tradert og stilltiende overenskomst språkbrukerne imellom», og som dermed «muliggjør og regulerer den faktiske språkbruken». Denne typen operative subnormer av bokmål har i faglitteraturen vore kalla *konservativt*, *moderat* og/eller *radikalt* bokmål.

2.2 Nemningane *konservativt*, *moderat* og *radikalt* bokmål

I ein artikkel som drøftar dei tre nemningane meir inngående, skriv Rauset (2022: 226):

Ordskitet både i tidlegare tider og nyare publikasjonar viser at nemningane *konservativt*, *moderat* og *radikalt bokmål* ikkje er veldefinerte og uproblematisk; vi har sett at det varierer kva folk legg i nemningane, og om ein opererer med ei todeling, tredeling eller fleire subvarietetar av bokmålet. For somme handlar dikotomien *konservativt* versus *radikalt bokmål* om mindre og større endringar i norma i eit språkhistorisk perspektiv, medan skiljet mellom *moderat* versus *radikalt bokmål* for andre handlar om stilistiske forskjellar og at dei radikale formene er tilnæringsformer til nynorsk og/eller folkemålet på Austlandet. Frå riksmålshald har vi sett døme på at moderat

og radikalt bokmål blir definerte ut frå avstand til riksmålet. Mange brukar nemninga *moderat* om konservative former, det gjeld ikkje minst dei som nyttar denne subvarieteteten.

Det er liten tvil om at dei tre nemningane har vore og blir brukte på skiftande vis, noko Rauset (2022) også dokumenterer ved å vise til ulike oppgåver, avhandlingar og artiklar som omhandlar tematikken (sjå til dømes Vinje 2004; Dyvik 2012; Kola 2014; Fjeld 2015; Helset 2017), noko som inneber at det substansielle innhaldet bak nemningane på ingen måte kan seiast å vere eintydig. I denne artikkelen vil vi ikkje gå i djupna av ordskiftet om dei tre nemningane, men derimot leggje vekt på at kvar av dei tre nemningane ikkje berre kan bli brukt om einskildvariantar, men også om varietetar. For sjølv om alle som skriv bokmål, i prinsippet kan velje fritt på kryss og tvers mellom eit stort tal tillatne og no sidestilte variantar innanfor den offisielle norma, slår *Retningslinjer for normering av bokmål og nynorsk* altså fast at tekstmaterialet av moderne bokmål «viser ein del samvariasjon mellom dei ulike formvala, slik at vi kan snakka om operative subnormer av bokmål i form av varietetar som tradisjonelt har vore kalla «konservativt», «moderat» og «radikalt» bokmål, med mellomformer» (Språkrådet 2021: 11). Det er nemleg nettopp kartlegging av normklynger av bestemte språklege formval som kan la seg identifisere som meir eller mindre dominerande operative subnormer av bokmål, som står i søkjelyset i denne studien. Ikkje dess mindre måtte vi ta stilling til om vi skulle operere med to eller fleire subvarietetar og dessutan kva for nokre spesifikke variantar eller ordformer vi skulle velje ut som representantar for dei ulike varietetane. Vi bestemte oss for å operere med ei todeling og bruke dei to nemningane *moderat* og *radikalt* bokmål, då dette har vore mest vanleg i faglitteraturen (Rauset 2022: 217). Med bakgrunn i det nemnde ordskiftet definerte vi *moderat bokmål* som *(eit sett av) tradisjonelle eller konservative former som maksimalt inneber moderate endringar i høve til 1917-normalen, og som ligg innanfor 1981-normalen eller 2005-normalen*, medan vi definerte *radikalt bokmål* som *(eit sett av) tilnæringsformer til nynorsk og/eller folkemålet på Austlandet som inneber radikale endringar i høve til 1917-normalen, og som ligg innanfor 1981-normalen eller 2005-normalen*. Sjølv om vi la til grunn desse operasjonelle definisjonane av høvesvis moderat og radikalt bokmål, for slik å lettare kunne velje ut dei språklege variablane som skulle vere med i studien, er det grunn til å peike på at det truleg er slik

at nokre av dei utvalde radikale formene vil bli oppfatta som meir radikale enn andre, på same måten som nokre av dei utvalde moderate formene vil bli oppfatta som meir tradisjonelle eller konservative (og dermed mindre moderate) enn andre.

2.3 *Tidlegare studiar av bokmålstekstar*

Det har vore gjennomført ei rekkje studiar av bruken av konservative, moderate og radikale former i bokmål. Typisk for dei fleste studiane som tek føre seg faktisk skriftleg språkbruk av bokmål, er at dei kartlegg eit avgrensa tal språklege variablar innanfor éin type tekstar (Kola 2014: 8ff). Når det gjeld avistekstar, har Vinde (2000) granska bruken av utvalde variablar i eit stort utval aviser og i NTB, medan Roksvold (2005) har granska utviklinga i eit utval aviser i eit diakront perspektiv og Goffeney (2013) har granska både rapportert og faktisk språkbruk mellom journalistar i ei bestemt avis. Sjølv om desse studiane har ulike innfallsvinklar og studieobjekt, trekkjer dei alle den same overordna konklusjonen, nemleg at moderate former dominerer i avistekstar. Mykje av det same kan seiast om studiar som er gjorde av romanar og sakprosatetekstar. I ein studie av former i romanar frå perioden 1937–1977 finn Pettersen (1993: 163) at fleirtalet av forfattarane «holder seg til det vi kan kalle en moderat bokmålsform», medan Baardsgaard (2005: 95) i ein studie av romanar frå perioden 1974–2001 konkluderer med at utviklinga har gått mot ein normal som er prega av moderate former. Mosling (1998: 56), som har granska formvariasjonen i åtte lærebøker i UH-sektoren, konkluderer på same måten: «De moderate formene er så å si enerådende.»

To studiar skil seg ut frå dei ovanfor nemnde studiane ved at dei for det første legg til grunn større og meir samansette utval av tekstar og for det andre kartlegg bruken av ulike språklege variablar ved hjelp av søk i elektroniske korpus. I ein studie av Rosén (2000) som særleg la vekt på å granske bruken av (tidlegare) sideformer i bokmål, blei det gjort søk i Oslo-korpuset, som består av tagga tekstar med til saman 18,5 millionar ord frå aviser, vekeblad, romanar, NOU-rapportar, lover og forskrifter. Studien konkluderer med at dei moderate formene dominerer, medan ei lang rekkje tillatne radikale sideformer er svært lite brukte. Den mest omfattande studien av formval i bokmål som er gjennomført til no, er masteroppgåva til Kjersti W. Kola (2014). Kola (2014: 51) har kartlagt formvariasjonen for til saman 92 morfologiske og ortografiske variablar med valfridom innanfor Leksikografisk bokmålskorpus, som består av

eit breitt samansett utval av tagga tekstar innanfor ei rekkje ulike sjangrar med i overkant av 100 millionar ord til saman. Studien til Kola (2014: 57) syner at dei konservative variantane dominerer over dei radikale motsvara deira i heile 84 % av tilfella, medan dei radikale variantane berre dominerer i 16 % av tilfella.²

Ein studie av Dyvik (2012) skil seg ut frå alle dei ovanfor nemnde studiane ved at han granskar *samanhengar mellom formval* i eit utval av språklege variablar på tvers av eit utval tekstar. Ved hjelp av statistisk korrespondanseanalyse undersøker Dyvik (2012: 218) førekomsten av normklynger innanfor eit aviskorpus beståande av bokmålstekstar, og han finn indikasjonar på eksistensen av ei klynge beståande av moderate former. Men som Dyvik (2012: 218) sjølv strekar under, var hovudsiktemålet med studien å teste ut metoden i seg sjølv: Både utvalet av tekstar og utvalet av språklege variablar i studien var for lite og for smalt til at ein kunne trekkje konklusjonar om operative normer i bokmål. Som ein ser, er dei fleste av dei tidlegare studiane av bokmål anten gamle eller utførte innanfor éin type tekstar, samstundes som berre éin av dei granskar samanhengar mellom formval på tvers av ulike variablar. Det er derfor på sin plass med ein ny studie av eit breitt og samansett utval av nyare bokmålstekstar.

3 Materiale og metode

Proposisjonen for språklova og dei nemnde retningslinjene for normering av bokmål slår altså fast at eventuelle justeringar av bokmålsrettskrivinga bør byggje «på undersøkingar av store korpus av bokmålstekstar som har god sjangerspreiing og regional spreining, og som både speglar dagens språk og går nokre tiår tilbake i tid» (Kulturdepartementet 2019–2020; Språkrådet 2021). På denne bakgrunnen, og fordi studiar av digitaliserte og tagga korpus set oss i stand til å avdekkje bruksfrekvensar og samvariasjon på tvers av store mengder tekst på effektive måtar, har vi i denne studien valt ei korpuslingvistisk tilnærming. I ka-

2. Kola (2014: 7) har valt å bruke nemningsparet «konservative» vs. «radikale» former, i staden for «moderate» vs. «radikale» former, og definerer konservative former som «de mest tradisjonelle, de som har vært lengst i bokmålsnormalen», medan ho definerer radikale former som «de nyere variantene, de som har kommet inn seinere, enten fra nynorsk eller fra dialektene.»

pittel 3.1 grunngjev vi kvifor vi har valt nettopp LBK som grunnlag for studien, før vi i kapittel 3.2 gjer greie for kva for språklege variablar vi har valt å granske førekomstar og samvariasjon av. McEnergy og Hardie (2012: 5) trekkjer opp eit skilje mellom korpusbaserte og korpusdrivne studiar, der førstnemnde er kjenneteikna av at ein brukar korpus til å teste førehandsutvikla teoriar eller hypotesar, medan sistnemnde er kjenneteikna av at ein brukar sjølvne korpuset til å utvikle hypotesar om språket. I denne studien har vi brukt ei korpusbasert, hypotetisk-deduktiv tilnærming og dessutan statistisk korrespondanseanalyse, noko som vil bli gjort nærmare greie for i kapittel 3.3.

3.1 Val av korpus

McEnergy og Hardie (2012: 2) poengterer at det er avgjerande viktig at korpuset som ein legg til grunn for ein studie, er veltilpassa forskings-spørsmåla for studien. Sidan det overordna forskings-spørsmålet for denne studien er om det eksisterer normklynger som kan la seg identifisere som meir eller mindre dominerande operative subnormer innanfor moderne bokmål, vil det vere føremålstenleg å legge til grunn eit korpus som både er balansert og representativt for dette skriftspråket. McEnergy og Hardie (2012) definerer «balanse» og «representativitet» slik:

A corpus is said to be *balanced* if the relative sizes of each of its subsections have been chosen with the aim of adequately representing the range of language that exists in the population of texts being sampled (s. 239).

A *representative* corpus is one sampled in such a way that it contains all the types of text, in the correct proportions, that are needed to make the contents of the corpus an accurate reflection of the whole of the language or variety that it samples (s. 250).

LBK er nettopp eit slikt balansert og representativt korpus som er spesielt tilrettelagt for utforsking av leksikalsk, ortografisk og morfologisk variasjon i moderne norsk bokmål, og det består av om lag 27 000 tekstar henta frå perioden 1985–2013 og inneheld i overkant av 100 millionar ord (Fjeld et al. 2020: 101, 107; Tekstlaboratoriet 2022). Derfor valde vi dette korpuset som utgangspunkt for denne studien. Clarino Centre Bergen har fått bruke ein versjon av LBK i sin Corpuscle, og det er denne versjonen som ligg til grunn for vår studie. Fordi den statistiske metoden vi ville bruke i studien, føreset at vi kan skilje formvala i einskildtekstane

frå kvarandre, valde vi å legge til grunn eit bestemt utval av tekstane i LBK når vi skulle kartlegge talet på førekomstar av moderate og radikale former innanfor kvar av dei ortografiske og morfologiske variablane som vi ville granske. Det blei lagt vekt på at subkorpuset skulle vere balansert og representativt i høve til alle tekstane i LBK, vel å merke med unntak av dei uredigerte tekstane i LBK.³ Samstundes stilte vi som premiss at kvar tekst som skulle vere med i subkorpuset, måtte innehalde minimum 15 000 ord, då med eit unntak for einskildartiklar. Subkorpuset som vi sette saman, bestod av til saman 452 tekstar og i alt 24 883 794 ord (dvs. om lag ¼ av det samla ordtilfanget i heile LBK), og tekstane fordelte seg slik:

- Periodika (5,8 %): Tekstutdrag frå 5 riksdekkande aviser, 3 region-aviser og 6 vekeblad
- TV (4,5 %): Teksting av 13 nyhende- og dokumentarprogram og 8 seriar og filmar
- Skjønnlitteratur (36,0 %): 40 romanar, 24 novellesamlingar, 3 drama og Bibelen
- Sakprosa (53,7 %): 147 einskildartiklar, 20 lærebøker, 7 fagbøker, 8 biografiar, 30 avhandlingar, 30 juridiske dokument, 30 rapportar, 9 reiseskildringar og utdrag frå 2 leksikon⁴

Det blei lagt vekt på at fordelinga mellom tekstar som er publiserte i det tidsrommet då bokmålsnormalen av 1981 var gjeldande og tekstar som er publiserte i det tidsrommet då bokmålsnormalen av 2005 var gjeldande, skulle vere så jamn som råd. Såleis er 54,3 % av tekstane i subkorpuset publiserte i tidsrommet 1985–2004, medan 45,7 % av tekstane er

3. Hovudgrunnen til at vi har valt å berre legge redigerte tekstar til grunn for studien, er at dei offisielle retningslinjene for normering av bokmål slår fast at normeringsgrunnlag i første rekke *skal* vere tekstar som har vore gjennom ein redigeringsprosess (Språkrådet 2021: 12). Dessutan er det slik at sjølv om ein del avisredaksjonar og forlag kan ha eigne husnormer og retteprogram som er snevrare enn den offisielle norma for bokmål, representerer denne typen redigerte tekstar den typen bokmål som norske språkbrukarar i størst grad de facto møter i det daglege livet, noko som i sin tur fører til at dei dannar hovudgrunnlaget for dei operative normene i språksamfunnet.
4. Til samanlikning er fordelinga i heile LBK slik: 5% periodika, 5 % tv-tekst, 35 % skjønnlitteratur, 50% sakprosa og 5 % upublisert og uredigerte tekstar, noko som vitnar om at subkorpuset er balansert og representativt for LBK.

publiserte i tidsrommet 2005–2013. Det var også ønskeleg med størst moglege balanse mellom kjønn på forfattarane i subkorpuset. Dette var likevel ikkje så lett å få til, sidan berre 30 % av forfattarane av tekstane i LBK er kvinner, medan 53 % er menn, samstundes som ein manglar opplysningar om kjønn på forfattar for resten av tekstane (Fjeld et al. 2020: 111). Såleis enda vi opp med 9,9 % kvinnelege forfattarar og 19,2 % mannlige forfattarar i subkorpuset, medan vi manglar opplysningar om kjønn på forfattar for storparten av tekstane.

Av lett forståelege årsaker er det etter måten stor variasjon i talet på ord innanfor kvar av tekstane, slik at nokre tekstar vil telje meir enn andre, både når det gjeld absolutte frekvensar og når det kjem til lokalisering av eventuelle normklynger, noko som er uunngåeleg dersom ein skal femne om eit breitt spekter av ulike typar tekstar. Likevel er det altså slik at utanom dei 147 einskildartiklane består kvar av dei 305 resterande tekstane i materialet av minimum 15 000 ord, og dei fleste av desse att består av meir enn 30 000 ord. På dette grunnlaget vil vi hevde at data-materialet som ligg til grunn for studien, samla sett gjev eit nokså balansert og representativt bilete av språklege val innanfor redigerte og publiserte tekstar på moderne bokmål, og dermed er heilt tråd med dei nye offisielle retningslinjene for normering av bokmål, der det altså heiter at «[d]et vesentlege grunnlaget for normeringa av bokmål er språket slik det ser ut i bokmålstekstar», og at normeringsgrunnlaget i første rekke skal vere «tekstar som har vore gjennom ein redigeringsprosess» (Språkrådet 2021: 12).

3.2 *Val av språklege variablar*

Det har vore stor og skiftande valfridom innanfor bokmål gjennom dei siste hundre åra, og ifølgje Fjeld (2015: 37) er valfridomen faktisk større i gjeldande bokmålsnormal enn i gjeldande nynorsknormal. Sidan tekstane i LBK var produserte mellom 1985 og 2013, valde vi å rette søkjelyset mot eit utval av språklege variablar som hadde to eller fleire alternative skrivemåtar innanfor 1981-normalen eller 2005-normalen, der den eine eller andre varianten kan bli rekna for å vere anten moderat eller radikal. Vi såg det som føremålstenleg å få med eit breitt og samansett utval av variablar, og la stor vekt på å få med variablar som representerte alle moglege typar av valfridom både innanfor staving og innanfor bøyging. Med dette som utgangspunkt og med bakgrunn i dei ovanfor nemnde definisjonane av høvesvis moderat og radikalt bokmål, valde vi ut til saman 150

språklege variablar som vi ville kartlegge førekomstane av i subkorpuset. 50 av variablane gjeld staving, medan 100 gjeld bøyning.

Av dei 50 ortografiske variablane er det 30 som gjeld vokalisme og 12 som gjeld konsonantisme, medan åtte av variablane involverer både vokalar og konsonantar. Som døme på variablar der det er valfridom innanfor vokalisme, kan vi nemne *fram/frem* (adverb). Som døme på ein variabel med valfridom innanfor konsonantisme kan vi trekkje fram *grøn/grønn* (adjektiv), medan vi som døme på valfridom som involverer både vokalar og konsonantar, kan nemne *skau/skog* (substantiv). Av dei 100 morfologiske variablane er det 77 som gjeld ulike typar substantivbøyning og 23 som gjeld ulike typar verbbøyning. Som døme på variablar innanfor substantivbøyninga kan vi nemne *døra/døren* (substantiv i bunden form eintal med valfri ending på *a/en*), *foreldra/foreldrene* (hankjønns-substantiv i bunden form fleirtal med valfri ending på *a/ene*), og *brev/brevene* (inkjekjønns-substantiv i bunden form fleirtal med valfri ending på *a/ene*). Som døme på variablar innanfor verbbøyninga kan vi trekkje fram *snakka/snakket* (verb i preteritum med valfri ending *a/et*) og *skreiv/skrev* (sterke verb med valfri bøyning i preteritum). I alle desse døma er det som kan bli rekna som det meir eller mindre radikale alternativet sett opp til venstre, medan det som kan bli rekna som det meir eller mindre moderate alternativet, er sett opp til høgre – og det same gjeld i Tabell 1–6 i resultatkapittelet. Som allereie nemnt i 2.1 er det gjerne slik at nokre av dei utvalde radikale formene vil bli oppfatta som meir radikale enn andre, på same måten som nokre av dei utvalde moderate formene vil bli oppfatta som meir tradisjonelle eller konservative (og dermed mindre moderate) enn andre. Til dømes tilseier vår eigen intuisjon at ein variant som *skau* av dei fleste bokmålsbrukarar vil bli rekna som meir radikal enn ein variant som *fram*, medan ein variant som *døren* moglegvis vil bli rekna som meir konservativ enn ein variant som *foreldrene*. Ein kan tenke seg at dette gjev seg utslag i talet på førekomstar av dei ulike variantane i subkorpuset på den måten at ein finn færre førekomstar av det som blir rekna som svært radikale og svært konservative former, enn det som blir rekna for mindre radikale og mindre konservative former. Derfor la vi vekt på å få med variablar som inneheld variantar som er alt frå det som vår eigen intuisjon tilseier er svært radikale former, til variantar som intuisjonen tilseier er svært konservative former i det samla variabelutvalet for studien. Likevel valde vi å setje ei nedre grense på minimum 100 førekomstar av den eine eller den andre varianten i

subkorpuset for at ein variabel skulle kome med i presentasjonen i denne artikkelen. Hovudgrunnen til det er at det knyter seg for stor statistisk uvisse til variablar med få førekomstar. Sidan utvalet av dei 150 variablane som vi skulle kartlegge førekomstane av, blei gjort *før* vi starta sjølve kartlegginga i subkorpuset, inneheld tabellane som er presenterte i resultatkapittelet, dermed samla sett noko færre enn 150 variablar.

McEnery og Hardie (2012: 13) trekkjer opp eit skilje mellom korpus som er tagga, og korpus som ikkje er det. I eit ikkje-tagga korpus kan ein berre søkje på einskildord, frasar eller setningar, utan omsyn til kva for type det er tale om, medan det i eit tagga korpus vil vere slik at ulike språklege einingar er merkte på eit eller anna vis. For vårt føremål var det avgjerande viktig at LBK var tagga med oppslagsord, ordklasse og bøyingskategoriar ved hjelp av *Oslo-Bergen-taggeren*. På denne bakgrunnen kunne dataekspertisen ved Clarino Bergen Centre hente ut ei matrise som synte talet på førekomstar av kvar av dei valfrie variantane av dei 150 utvalde variablane innanfor kvar av dei 452 tekstane i subkorpuset vi hadde sett saman. Ei gjeven celle i matrisa syner altså kor mange førekomstar ein finn av den aktuelle varianten i den aktuelle teksten, medan resultatcellene i botnen av kvar kolonne syner kor mange førekomstar av varianten ein finn i alle tekstane til saman.

3.3 Statistisk korrespondanseanalyse

Det empiriske utgangspunktet for studien var altså eit større utval av språklege variablar med valfridom og eit stort, balansert og representativt korpus bestående av bokmålstekstar, og føremålet med studien var å finne eventuelle klyngestrukturar i dette tekstutvalet. Intuisjonen er at valfridomen i bokmålsnorma skaper eit rom av moglegeheiter, og kvart innbyrdes uavhengige sett av formalternativ kan bli sett på som ein dimensjon i dette rommet. Det som då plasserer seg i dette rommet, og kanskje dannar klynger, er tekstar, medan sett av språklege formval som er typiske for ei klynge av tekstar, vil kunne bli definerte som normklynger. På denne måten set metoden oss i stand til avdekkje og synleggjere komplekse system av relasjonar innanfor store datamatriser ved hjelp av geometriske representasjonar som syner kva for tekstar som har ein tendens til å innehalde dei same språklege formvala, og kva for formval som har ein tendens til å opptre i lag på tvers av tekstane i korpuset. Derfor er korrespondanseanalyse ein sær sars eigna metode til å skaffe eit kunnskapsgrunnlag for skriftspråksnormering (Dyvik 2012; Helset 2017).

Metoden er induktiv og utforskande på det viset at ein freistar å avdekkje mønster på bakgrunn av eit empirisk materiale. Ein tek utgangspunkt i språket slik det ligg føre i faktiske tekstar med tanke på å avdekkje skriftspråkets operative normer. Input er ei matrise med tekstar langs den eine dimensjonen og språklege variablar langs den andre som syner korleis dei valfrie variantane av dei utvalde språklege variablane fordeler seg i det samla tekstutvalet og korleis tekstane fordeler seg med omsyn til variantane innanfor ulike språklege variablar. Eit vesentleg poeng i denne samanhengen er at ein ikkje a priori fastset dei dimensjonane som er utgjorde av bestemte sett av alternative former. Dei faktiske dimensjonane i rommet blir rekna ut via korrespondanseanalysen sjølv, ut frå eit føremål om å gjere greie for avstandane mellom tekstar og formval ved hjelp av eitt eller fleire todimensjonale koordinatsystem, der nokre av dei viktigaste språklege skilnadene blir framstilte langs den eine aksens, medan andre viktige språklege skilnader blir framstilte langs den andre aksens.

Sjølv om grunnleggjaren av metoden (Benzécri 1969, 1973) la vekt på å utvikle ein induktiv og utforskande teknikk for analyse av lingvistiske data, understrekar Le Roux og Rouanet (2010: 11) at dersom metoden skal vere fruktbar, må ein bruke relevante data: «Performing a geometric analysis does *not* mean gathering disparate data and looking for «what comes out» of the computer». McEnery og Hardie (2012: 161) understrekar også dette poenget: «Put bluntly, no one approaches corpus data with no preconceptions whatever (.....) All analysts bring some prior understanding of language to the analysis; this is unavoidable and, in all likelihood, useful.» Derfor valde vi som forklart i kapittel 3.2 å granske eit utval på 150 språklege variablar som hadde to eller fleire alternative skrivemåtar innanfor 1981-normalen eller 2005-normalen, der den eine eller andre varianten kan bli rekna for å vere anten moderat eller radikal, slik desse nemningane er defnerte i teorikapittelet ovanfor.

Den berande teknikken innanfor korrespondanseanalyse byggjer på logikken som er kjent frå klassisk kjikvadrat-analyse av krysstabellar. Kjikvadratet (χ^2) finn vi ved å summere det kvadrerte avviket mellom forventa (normalfordelte) og observerte verdiar dividert med forventa verdiar for kvar av cellene i radene og kolonnane i ein gjeven tabell. Eit avgjerande punkt er om χ^2 er stort nok i høve til talet på fridomsgrader i tabellen (som ein finn ved å multiplisere (talet på rader – 1) med (talet på kolonnar – 1)) til at det rår ein statistisk samanheng mellom rad- og ko-

lonnevariablane i tabellen. Dette finn ein ut av ved χ^2 -testen, der nullhypotesen (H_0) er at det *ikkje* er ein statistisk samanheng (Hjellbrekke 1999: 22; Greenacre 2007: 27). Dersom χ^2 er over kritisk verdi, kan vi forkaste H_0 og gå vidare med H_1 , som postulerer at avviket mellom forventa og observerte verdier er så stor at det *er* ein statistisk samanheng mellom rad- og kolonnevariablane. I så tilfelle er det grunnlag for å gå vidare med analysane og konstruere eit mangedimensjonalt og vekta koordinatsystem på bakgrunn av verdiane i tabellen. Talet på rader (R) og kolonnar (K) i krysstabellen avgjer talet på dimensjonar i koordinatsystemet, som alltid vil vere det minste av $(R - 1)$ og $(K - 1)$ (Hjellbrekke 1999: 30). Dei språklege variantane i kolonnevariablane og tekstane i radvariablane kan framstillast som to punktskyer i det same mangedimensjonale rommet, der kvar variant og kvar tekst let seg lokalisere som punkt ut frå avstanden til barysenteret, som er det vekta gjennomsnittet av alle variantane/tekstane, og ut frå avstanden til kvarandre innanfor kvar av dei to punktskyene (Lebart 1998: 54; Hjellbrekke 1999: 34; Greenacre 2007: 17).

Variansen i ei punktsky blir kalla inertia (φ^2). Vi finn den totale variansen i ei punktsky ved å rekne ut summen av massen multiplisert med den kvadrerte avstanden til barysenteret for kvart enkelt punkt i skya, eller ved å dividere kjikvadratet på det samla talet på førekomstar i materialet: $\varphi^2 = \chi^2/n$ (Greenacre 2007: 28). Høg varians for ei sky impliserer stor spreining av punkta i skya, noko som igjen impliserer stabilitet, dvs. at det er høg statistisk samanheng mellom verdiane i materialet. Avstanden til barysenteret og massen til eit punkt (dvs. ein språkleg variant eller ein tekst) er avgjerande for kor mykje punktet bidreg til den totale variansen. Dess større masse og dess større avstand, dess høgere bidrag (Hjellbrekke 1999: 37).

Korrespondanseanalysen gjev oss på dette viset eit mangedimensjonalt rom bestående av to punktskyer der kvart punkt står i ein gjeven distanse til barysenteret. Men sidan det er vanskeleg for det menneskelege sinnet å førestille seg slike mangedimensjonale rom, vil ein innanfor korrespondanseanalysen freiste å avdekkje den underliggende informasjonen i ei matrise ved å framstille den viktigaste informasjonen ved hjelp av eitt eller fleire todimensjonale koordinatsystem, der nokre av dei viktigaste språklege skilnadene blir framstilte langs den eine aksene og andre viktige språklege skilnader blir framstilte langs den andre aksene. Som Hjellbrekke (1999: 34) skriv, krev dette at vi kan finne dei aksane

som dannar dei faktorplana som best kan skildre strukturane i punkt-skyene. Til kvar akse knyter det seg ein eigenverdi (λ) som syner kor mykje av den samla variansen han fangar opp. Summen av alle aksane sin eigenverdi ($\sum \lambda$) vil vere identisk med verdien til den totale variansen for skya (Hjellbrekke 1999: 37). I ein analyse vil det vere slik at den første aksan fangar opp dei mest framtrêdande motsetningane i materialet, og dermed den største prosentdelen av den totale variansen. Den andre aksan fangar opp (delar av) dei motsetningane som ikkje blei fanga opp av førsteaksen, medan den tredje aksan fangar opp (delar av) dei motsetningane som ikkje blei fanga opp av førsteaksen og andreaksen og så bortetter. På dette viset kan ein framstille den viktigaste informasjonen ved å tolke ulike todimensjonale koordinatsystem eller faktorplan (Hjellbrekke 1999: 35). Det finst ikkje noko fasitsvar på kor mange aksar og faktorplan ein skal tolke, men målet er å gå vidare med færrest moglege aksar samstundes som ein tek vare på mest moglege informasjon frå det opphavlege mangedimensjonale rommet/den opphavelge matrisa. Eit mykje brukt kriterium er det såkalla olbogekriteriet, som baserer seg på at ein først teiknar ei kurve som syner kor stor eigenverdi som knyter seg til dei ulike aksane, og som tilseier at ein skal ta med aksar til og med det punktet på kurva der ho tek til å flate ut (Hjellbrekke 1999: 38).

Før ein kan ta til å tolke resultatata, treng ein som både Lebart (1998: 58), Hjellbrekke (1999: 41) og Greenacre (2007: 81) understrekar, også å finne ut kva for språklege variantar og tekstar som har vore dei viktigaste for konstruksjonen av aksane i planet, og kor godt desse aksane er i stand til å skildre kvart einskilt punkt i det same planet. Det første finn ein ut ved å rekne ut dei absolutte bidraga, som syner bidraget frå eit punkt til ein bestemt akse. Det andre finn ein ved å rekne ut dei relative bidraga, som syner kor mykje av eit punkt som blir forklart av ein bestemt akse. Vanlegvis er det slik at punkt som yter høge absolutte bidrag til ein akse, også mottek høge relative bidrag frå den same aksan, men ikkje naudsynlegvis omvendt (Hjellbrekke 1999: 44).

Sidan den offisielle skriftspråksnorma som var gjeldande i den perioden tekstane i LBK blei til i, ikkje seier noko om korleis ein bør kombinere ulike ordformer innanfor ein tekst, føreslår norma ein nullhypotese (H_0) om at variantane og tekstane vil spreie seg jamt utover i rommet, dvs. ei tilnærma normalfordeling med $\chi^2 \approx 0$. Men sidan den tidlegare nemnde studien til Dyvik (2012: 218) indikerer eksistensen av operative normer, dvs. «en tradert og stilltiende overenskomst språkbru-

kerne imellom» med nokså klare førestillingar om kva for former som høyrer saman med kvarandre, kunne vi forvente opphoping av variantar og tekstar i visse delar av rommet også i denne studien. Derfor sette vi fram ein hypotese om at det innanfor det totale tekstutvalet ville vekse fram normklynger av bestemte ordformer som kunne la seg identifisere som statistisk signifikante subvarietetar av bokmål.

Innanfor korrespondanseanalyse er det som Hjellbrekke (1999: 36) skriv, slik at tekstar som inneheld (mange av) dei same variantane, og variantar som har ein tendens til å opptre i (mange av) dei same tekstane, hamnar nær kvarandre i rommet, medan tekstar og variantar som er gjensidig ekskluderande, hamnar langt frå kvarandre i rommet. Dei mest representative og minst spesielle tekstane og variantane hamnar nær barysenteret, medan dei mindre representative og meir spesielle tekstane og variantane hamnar lengre ute i periferien. Slik kan ein finne ordformer og kombinasjonar av ordformer som er meir brukte enn andre, og ikkje minst normklynger, som vi har definert som eit sett av språklege formval som er felles for eit antal tekstar på ein slik måte at tekstane skil seg som gruppe frå andre tekstar (Helset 2017).

4 Resultat og drøfting

I den første delen av dette kapittelet presenterer vi tabellar som syner absolute og relative frekvensar av radikale og moderate former i subkorpuset av LBK. Vi startar med ei jamføring av talet på førekomstar av radikale og moderate variantar i eit utval av dei 50 ortografiske variablane, før vi jamfører førekomstar av radikale og moderate variantar i eit utval av dei 100 morfologiske variablane som er med i studien. Presentasjonen av absolute og relative frekvensar i det samla tekstutvalet vil gje eit godt grunnlag for å seie noko om kva som er gjeldande operative normer innanfor moderne bokmål – og kan slik vere utfyllande i høve til tidlegare studiar på feltet. Dette gjeld vel å merke berre fordeling mellom ulike variantar innanfor kvar av dei einskilde språklege variablane som er med i utvalet. For å kartlegge eventuelle operative normklynger beståande av fleire ulike språklege variantar på tvers av tekstane i subkorpuset, har vi altså brukt metoden statistisk korrespondanseanalyse, og i den andre delen av dette kapittelet vil vi presentere slike analysar. I forlenginga av dette peikar vi på kva for implikasjonar funna i studien har for det fram-

tidige normeringsarbeidet for bokmål, gjeve dei offisielle retningslinjene for normering av bokmål og nynorsk som vi presenterte i 1 ovanfor (Språkrådet 2021).

4.1 Ortografiske variablar

Som nemnt i metodekapittelet valde vi å setje ei nedre grense på minimum 100 førekomstar av den eine eller den andre varianten for at ein variabel skal kome med i presentasjonen i denne artikkelen. Tabell 1 nedanfor syner absolutte og relative frekvensar av dei meir eller mindre radikale og moderate formene av 42 av dei 50 utvalde ortografiske variablane som synt seg å ha minst 100 førekomstar av den eine eller andre varianten i subkorpuset.

Tabellen avdekkjer at den moderate varianten er brukt i meir enn 99 % av tilfella for nøyaktig halvparten (50,0 %) av dei 42 ortografiske variablane og i mellom 90 og 99 % av tilfella for nøyaktig 1/3 (33,3 %) av variablane. Berre ved sju av dei 42 kartlagde variablane (16,7 %) er den moderate varianten brukt i mindre enn 90 % av tilfella. Moderate former som *løse* (verb), *lav* (adjektiv), *fra* (preposisjon), *løs* (adjektiv), *uke* og *vei* (substantiv) dominerer fullstendig over dei meir radikale motsvara *løyse*, *låg*, *laus*, *veke* og *veg*, som alle utgjer mindre enn 1,0 % av førekomstane av den aktuelle variabelen. Dominansen til dei moderate for-

Tabell 1 Ortografiske variablar⁵

| Radikal | Frekvens | Relativ frekvens | Moderat | Frekvens | Relativ frekvens |
|---------|----------|------------------|---------|----------|------------------|
| helse | 0 | 0,0 % | hilse | 646 | 100,0 % |
| klyppe | 0 | 0,0 % | klippe | 251 | 100,0 % |
| løyse | 0 | 0,0 % | løse | 1673 | 100,0 % |
| låg | 4 | 0,1 % | lav | 5813 | 99,9 % |
| kvit | 4 | 0,1 % | hvit | 5425 | 99,9 % |
| kjelde | 2 | 0,1 % | kilde | 2693 | 99,9 % |
| sleppe | 5 | 0,2 % | slippe | 2807 | 99,8 % |
| frå | 197 | 0,2 % | fra | 106215 | 99,8 % |
| grøn | 4 | 0,2 % | grønn | 2081 | 99,8 % |
| sakne | 1 | 0,2 % | savne | 482 | 99,8 % |
| leike/s | 1 | 0,3 % | leke/s | 353 | 99,7 % |
| hamn | 2 | 0,3 % | havn | 678 | 99,7 % |
| glas | 7 | 0,3 % | glass | 2361 | 99,7 % |
| høg | 42 | 0,3 % | høy | 13962 | 99,7 % |

5. «/s» bak eit ord tyder at det er tale om eit substantiv, medan «/v» bak eit ord tyder at det er tale om eit verb.

| | | | | | |
|---------|------|--------|----------|-------|--------|
| laus | 6 | 0,3 % | løs | 1893 | 99,7 % |
| flyge | 3 | 0,4 % | fly | 708 | 99,6 % |
| nase | 6 | 0,5 % | nese | 1236 | 99,5 % |
| veke | 39 | 0,7 % | uke | 5585 | 99,3 % |
| veg | 98 | 0,8 % | vei | 11687 | 99,2 % |
| mjøøl | 3 | 0,9 % | mel | 331 | 99,1 % |
| oske | 3 | 0,9 % | aske | 316 | 99,1 % |
| leike/v | 9 | 1,5 % | leke/v | 602 | 98,5 % |
| sjøl | 694 | 1,8 % | selv | 38714 | 98,2 % |
| skau | 42 | 1,8 % | skog | 2307 | 98,2 % |
| lauv | 5 | 2,3 % | løv | 210 | 97,7 % |
| tjuv | 8 | 2,9 % | tyv | 272 | 97,1 % |
| innafor | 204 | 3,7 % | innenfor | 5338 | 96,3 % |
| sein | 612 | 4,7 % | sen | 12331 | 95,3 % |
| gard | 70 | 5,1 % | gård | 1302 | 94,9 % |
| sjuk | 112 | 5,3 % | syk | 1995 | 94,7 % |
| dokke | 13 | 5,6 % | dukke | 221 | 94,4 % |
| gras | 65 | 7,7 % | gress | 775 | 92,3 % |
| heit | 31 | 8,7 % | het | 324 | 91,3 % |
| golv | 219 | 8,9 % | gulv | 2231 | 91,1 % |
| skeiv | 45 | 9,7 % | skjev | 417 | 90,3 % |
| bru | 134 | 12,6 % | bro | 928 | 87,4 % |
| likne | 262 | 21,5 % | ligne | 954 | 78,5 % |
| bein/s | 1360 | 44,2 % | ben/s | 1719 | 55,8 % |
| hogge | 112 | 47,1 % | hugge | 126 | 52,9 % |
| fram | 9920 | 51,3 % | frem | 9435 | 48,7 % |
| sju | 1229 | 60,0 % | syv | 820 | 40,0 % |
| verken | 1888 | 66,8 % | hverken | 938 | 33,2 % |

mene er også stor når det kjem til former som *selv* (determinativ), *tyv* (substantiv), *sen*, *syk* og *skjev* (adjektiv), der dei radikale motsvara *sjøl*, *tjuv*, *sein*, *sjuk* og *skeiv* utgjer mellom 1,0 % og 10,0 % av førekomstane av den aktuelle variabelen. Når det derimot gjeld variablane *bein/ben* (substantiv) og *hogge/hugge* (verb), ser vi at dei radikale variantane er nesten like mykje brukte som dei moderate, og når det gjeld variablane *fram/frem* (adverb), *sju/syv* (determinativ) og *verken/hverken* (konjunksjon) er det slik at det er fleire førekomstar av den radikale varianten enn av den moderate varianten i subkorpuset av LBK. Det siste kan ha samanheng med at variantane *fram*, *sju* og *verken* ikkje naudsynlegvis blir rekna som særleg radikale former, medan variantane *frem*, *syv* og *hverken* moglegvis blir rekna for å vere heller konservative former. Her vil det også vere relevant å sjå nærmare på rettskrivingshistorikken.

For 30 av desse 42 ortografiske variablane er det slik at dei to alternative variantane var og er sidestilte former i bokmål både i perioden 1981–2005 og i rettskrivinga av 2005, og for desse 30 kan ein såleis ikkje rekne med at den offisielle rettskrivinga har spela inn på bruken av den eine eller andre varianten. Når det derimot gjeld dei to variablane som har høgst del av det som her er sett opp som radikale former, *verken* (med 66,8 % av førekomstane) og *sju* (med 60,0 % av førekomstane), var dette dei einaste tillatne formene i perioden 1981–2005. Dei moderate motsvara *hverken* (med 33,2 % av førekomstane) og *syv* (med 40 % av førekomstane) blei først tillatne i og med bokmålsrettskrivinga av 2005. Vi kan vidare merke oss at høgfrekvente radikale former som *hogge* (med 47,1 % av førekomstane) og *golv* (med 8,9 %), var hovudformer i perioden 1981–2005, medan dei moderate motsvara *hugge* og *gulv* då var sideformer, før dei blei jamstilte i 2005. Det er såleis grunn til å tru at den offisielle rettskrivinga kan ha spela i favør av bruken av former som *verken*, *sju*, *hogge* og *golv* fram til 2005. Derimot kan dette ikkje vere ei årsaksforklaring når det kjem til det faktum at formene *fram* og *bein* er om lag like mykje brukte som formene *frem* og *ben*, då båe formene i desse to variablane har vore jamstilte sidan 1981. Det overordna biletet som teiknar seg, er iallfall at det ved dei aller fleste av dei 42 granska variablane er slik at dei moderate variantane dominerer temmeleg mykje over dei meir radikale motsvara. Samstundes har vi altså sett at fordelinga er meir jamn for variablane *bein/ben*, *hogge/hugge*, *fram/frem*, *sju/syv* og *verken/hverken*.

4.2 Morfologiske variablar

Av dei 100 morfologiske variablane som vi valde å kartlegge førekomstane av, er det som tidlegare nemnt 77 som gjeld ulike typar substantivbøying og 23 som gjeld ulike typar verbbøying, og i dette delkapittelet skal vi sjå nærmare på ulike undergrupper av desse i tur og orden.

4.2.1 Substantiv i bunden form eintal med valfri ending *a/en*

Mellom dei 77 variablane som gjeld substantivbøying, valde vi ut 44 variablar innanfor den store gruppa av substantiv som har valfri ending *a/en* i bunden form eintal. Av desse var det 30 variablar som stetta kravet om minst 100 førekomstar av den eine eller andre varianten i subkorpuset, og Tabell 2 nedanfor syner absolutte og relative frekvensar av dei meir eller mindre radikale og moderate variantane av desse variablane.

Tabell 2 Substantiv i bunden form eintal med valfri ending *a/en*

| Radikal | Frekvens | Relativ frekvens | Moderat | Frekvens | Relativ frekvens |
|------------|----------|------------------|-------------|----------|------------------|
| festninga | 0 | 0,0 % | festningen | 171 | 100,0 % |
| saka | 53 | 1,2 % | saken | 4492 | 98,8 % |
| spenninga | 8 | 1,5 % | spenningen | 534 | 98,5 % |
| utviklinga | 89 | 2,7 % | utviklingen | 3181 | 97,3 % |
| tegninga | 11 | 4,7 % | tegningen | 221 | 95,3 % |
| makta | 56 | 4,9 % | makten | 1092 | 95,1 % |
| hånda | 275 | 5,1 % | hånden | 5146 | 94,9 % |
| mora | 274 | 7,1 % | moren | 3595 | 92,9 % |
| tida | 829 | 9,0 % | tiden | 8406 | 91,0 % |
| mappa | 13 | 10,7 % | mappen | 108 | 89,3 % |
| natta | 336 | 11,8 % | natten | 2509 | 88,2 % |
| gruppa | 246 | 13,7 % | gruppen | 1548 | 86,3 % |
| skylda | 54 | 14,2 % | skylden | 326 | 85,8 % |
| avisa | 220 | 15,5 % | avisen | 1201 | 84,5 % |
| døra | 932 | 20,9 % | døren | 3536 | 79,1 % |
| gata | 398 | 23,9 % | gaten | 1270 | 76,1 % |
| kista | 61 | 24,2 % | kisten | 191 | 75,8 % |
| stua | 297 | 25,8 % | stuen | 854 | 74,2 % |
| elva | 449 | 27,9 % | elven | 1159 | 72,1 % |
| uka | 318 | 30,6 % | uken | 722 | 69,4 % |
| sola | 560 | 31,8 % | sølen | 1203 | 68,2 % |
| dyna | 76 | 32,5 % | dynen | 158 | 67,5 % |
| jorda | 1246 | 37,0 % | jorden | 2123 | 63,0 % |
| boka | 1169 | 37,3 % | boken | 1966 | 62,7 % |
| kona | 576 | 79,8 % | konen | 146 | 20,2 % |
| hytta | 383 | 84,5 % | hytten | 70 | 15,5 % |
| bygda | 205 | 89,9 % | bygden | 23 | 10,1 % |
| jenta | 1477 | 97,3 % | jenten | 41 | 2,7 % |
| kua | 124 | 98,4 % | kuen | 2 | 1,6 % |
| handa | 136 | 100,0 % | handen | 0 | 0,0 % |

I motsetnad til kva som er tilfellet med dei ortografiske variablane, der dei moderate variantane er fullstendig dominerande i dei aller fleste tilfella, avdekkjer Tabell 2 som heilskap det vi kan karakterisere som ein glidande overgang frå variablar der dei moderate variantane er temmeleg dominerande til variablar der dei radikale variantane er temmeleg dominerande. Tabellen syner at den moderate varianten er brukt i meir enn 90 % av tilfella for om lag $\frac{1}{3}$ (30,0 %) av dei 30 variablane og i mellom 60 og 90 % av tilfella for nøyaktig halvparten (50,0 %) av variablane. Samstundes ser vi at det for seks av dei 30 kartlagde variablane (20,0 %) er slik at den radikale varianten er brukt i meir enn 80 % av tilfella, og at

det for ti av variablane (33,3 %) er slik at den radikale varianten blir brukt i mellom 20 og 80 % av tilfella. Moderate former som *festningen, saken, spenningen, tegningen* og *makten* dominerer fullstendig over dei meir radikale motsvara *festninga, saka, spenninga, tegninga* og *makta*, som alle utgjer mindre enn 5,0 % av førekomstane av den aktuelle variabelen. Samstundes er dominansen til dei radikale formene temmeleg stor når det kjem til former som *kua, jenta, bygda* og *hytta*, der dei moderate motsvara *kuen, jenten, bygden* og *hytten* utgjer mindre enn 16 % av førekomstane av den aktuelle variabelen. Dette siste er heller ikkje særleg overraskande, sidan det her er tale om substantiv med konkret og kvardagsleg tyding som har hatt hovudform på *-a* i rettskrivinga heilt sidan 1938 og fram til 2005. Når det gjeld variablar som *stua/stuen, elva/elven, uka/uken, sola/solen, dyna/dynen, jorda/jorden* og *boka/boken*, ser vi derimot ei jamnare fordeling mellom radikale og moderate former i subkorpuset av LBK, då dei radikale formene utgjer mellom 25 og 38 % av førekomstane av desse variablane. Samla sett kan vi seie at det er ein tendens til at dei moderate formene dominerer over dei radikale også i eintal av substantiv med valfri ending *a/en*, sjølv om denne tendensen ikkje på nokon måte kan seiast å vere like tydeleg som det biletet som teikna seg når det gjaldt dei ortografiske variablane.

4.2.2 Hankjønnsst substantiv i bunden form fleirtal med valfri ending *a/ene*
Vi valde også å sjå nærmare på seks variablar som gjeld hankjønnsst substantiv i bunden form fleirtal som har valfri ending *a/ene*. Av desse var det tre variablar som stetta kravet om minst 100 førekomstar av den eine eller andre varianten i subkorpuset, og Tabell 3 nedanfor syner absolutte og relative frekvensar av dei meir eller mindre radikale og moderate variantane av desse variablane.

Tabell 3 Hankjønnsst substantiv i bunden form fleirtal med valfri ending *a/ene*

| Radikal | Frekvens | Relativ frekvens | Moderat | Frekvens | Relativ frekvens |
|----------|----------|------------------|------------|----------|------------------|
| skoa | 0 | 0,0 % | skoene | 494 | 100,0 % |
| foreldra | 19 | 0,9 % | foreldrene | 2056 | 99,1 % |
| tinga | 12 | 1,1 % | tingene | 1074 | 98,9 % |

Tabellen avdekkjer at den moderate varianten er brukt i meir enn 98 % av tilfella for alle dei tre variablane. Dei moderate formene *skoene, foreldrene* og *tingene* dominerer fullstendig over dei meir radikale motsvara

scoa, foreldra og *tinga*, som alle utgjer mindre enn 1,2 % av førekomstane av den aktuelle variabelen.

4.2.3 Inkjekjønnsst substantiv i bunden form fleirtal med valfri ending *a/ene*
Den siste typen substantiv vi såg nærmare på, var 27 variablar av den store gruppa av inkjekjønnsst substantiv som i bunden form fleirtal har valfri ending *a/ene*. Av desse var det 21 variablar som stetta kravet om minst 100 førekomstar av den eine eller andre varianten i subkorpuset, og Tabell 4 nedanfor syner absolutte og relative frekvensar av dei meir eller mindre radikale og moderate variantane av desse variablane.

Tabell 4 Inkjekjønnsst substantiv i bunden form fleirtal med valfri ending *a/ene*

| Radikal | Frekvens | Relativ frekvens | Moderat | Frekvens | Relativ frekvens |
|-----------|----------|------------------|-------------|----------|------------------|
| distrikta | 0 | 0,0 % | distriktene | 134 | 100,0 % |
| kontora | 0 | 0,0 % | kontorene | 107 | 100,0 % |
| menneska | 1 | 0,0 % | menneskene | 2111 | 100,0 % |
| bilda | 1 | 0,1 % | bildene | 1042 | 99,9 % |
| resultata | 1 | 0,1 % | resultatene | 1042 | 99,9 % |
| problema | 1 | 0,1 % | problemene | 893 | 99,9 % |
| landa | 6 | 0,4 % | landene | 1469 | 99,6 % |
| møta | 1 | 0,4 % | møtene | 244 | 99,6 % |
| husa | 3 | 0,5 % | husene | 572 | 99,5 % |
| språka | 1 | 0,6 % | språkene | 170 | 99,4 % |
| spørsmåla | 6 | 0,6 % | spørsmålene | 975 | 99,4 % |
| breva | 4 | 0,9 % | brevene | 442 | 99,1 % |
| behova | 2 | 1,2 % | behovene | 169 | 98,8 % |
| krava | 14 | 1,5 % | kravene | 920 | 98,5 % |
| fjella | 11 | 1,7 % | fjellene | 631 | 98,3 % |
| orda | 40 | 2,2 % | ordene | 1797 | 97,8 % |
| åra | 178 | 3,8 % | årene | 4455 | 96,2 % |
| dýra | 74 | 10,9 % | dyrene | 604 | 89,1 % |
| bena | 888 | 98,2 % | benene | 16 | 1,8 % |
| beina | 694 | 99,6 % | beinene | 3 | 0,4 % |
| barna | 3472 | 99,9 % | barnene | 2 | 0,1 % |

Tabellen avdekkjer at den moderate varianten er brukt i meir enn 99 % av tilfella for godt over halvparten (57,1 %) av dei 21 variablane og i mellom 89 og 99 % av tilfella for litt over 1/4 (28,6 %) av variablane. Dominansen til moderate former som *distriktene*, *kontorene*, *menneskene* etc. er med andre ord endå sterkare i denne variabelgruppa enn kva tilfellet var ved dei ortografiske variablane. Samstundes merker vi oss at dei radikale variantane *bena*, *beina* og *barna* fullstendig dominerer over dei moderate motsvara *benene*, *beinene* og *barnene*, som alle utgjer mindre

enn 2 % av førekomstane av desse variablane. Dette er iallfall ikkje overraskande for dei to sistnemnde variablane, då formene *beina* og *barna* er folkelege inkjekjønnsord med særnorsk lydform som har hatt hovudform på *-a* i rettskrivinga sidan 1938 og fram til 2005.

4.2.4 Verb i preteritum med valfri ending *a/et*

Mellom dei i alt 23 utvalde variablane som gjeld verbbøying, valde vi ut 16 variablar innanfor den store gruppa av verb som har valfri ending *a/et* i preteritum. Av desse var det 14 variablar som stetta kravet om minst 100 førekomstar av den eine eller andre varianten i subkorpuset, og Tabell 5 nedanfor syner absolutte og relative frekvensar av dei meir eller mindre radikale og moderate variantane av desse variablane.

Tabell 5 Verb i preteritum med valfri ending *a/et*

| Radikal | Frekvens | Relativ frekvens | Moderat | Frekvens | Relativ frekvens |
|----------|----------|------------------|-----------|----------|------------------|
| oppfatta | 4 | 0,8 % | oppfattet | 502 | 99,2 % |
| retta | 13 | 1,2 % | rettet | 1104 | 98,8 % |
| kasta | 26 | 1,2 % | kastet | 2083 | 98,8 % |
| hoppa | 7 | 1,4 % | hoppet | 498 | 98,6 % |
| ønska | 44 | 1,5 % | ønsket | 2928 | 98,5 % |
| åpna | 43 | 1,7 % | åpnet | 2449 | 98,3 % |
| slutta | 30 | 2,6 % | sluttet | 1142 | 97,4 % |
| snakka | 93 | 2,9 % | snakket | 3154 | 97,1 % |
| mista | 42 | 3,4 % | mistet | 1180 | 96,6 % |
| angra | 6 | 3,5 % | angret | 165 | 96,5 % |
| kutta | 4 | 3,7 % | kuttet | 103 | 96,3 % |
| mangla | 23 | 3,8 % | manglet | 582 | 96,2 % |
| vaska | 13 | 4,3 % | vasket | 286 | 95,7 % |
| jobba | 47 | 5,8 % | jobbet | 770 | 94,2 % |

Av tabellen ser vi at den moderate varianten er brukt i meir enn 94 % av tilfella for alle dei 14 variablane. Dominansen til moderate former som *oppfattet*, *rettet*, *kastet* etc. er altså meir overveldande i denne variabelgruppa enn kva tilfellet var med både dei ortografiske variablane og variablane som gjaldt bøying av inkjekjønnsst substantiv i bunden form eintal.

4.2.5 Sterke verb med valfri bøying i preteritum

Til slutt i denne delen skal vi sjå nærmare på nokre sterke verb med valfri bøying i preteritum. I dette tilfellet var det slik at alle dei sju variablane vi hadde valt ut, stetta kravet om minst 100 førekomstar av den eine eller andre varianten i subkorpuset, og Tabell 6 nedanfor syner ab-

solutte og relative frekvensar av dei meir eller mindre radikale og moderate variantane av desse variablane.

Tabell 6 Sterke verb med valfri bøying i preteritum

| Radikal | Frekvens | Relativ frekvens | Moderat | Frekvens | Relativ frekvens |
|---------|----------|------------------|---------|----------|------------------|
| gjekk | 19 | 0,1 % | gikk | 21773 | 99,9 % |
| fekk | 35 | 0,2 % | fikk | 22856 | 99,8 % |
| braut | 3 | 0,2 % | brøt | 1337 | 99,8 % |
| blei | 209 | 0,3 % | ble | 80067 | 99,7 % |
| flaut | 8 | 2,4 % | fløt | 322 | 97,6 % |
| skreiv | 98 | 2,5 % | skrev | 3849 | 97,5 % |
| drap | 35 | 3,8 % | drepte | 895 | 96,2 % |

Tabellen avdekkjer at den moderate varianten er brukt i meir enn 96 % av tilfella for alle dei sju variablane. Dei moderate preteritumsformene *gikk*, *fikk*, *brøt*, *ble*, *fløt*, *skrev* og *drepte* dominerer altså meir over dei radikale motsvara *gjekk*, *fekk*, *braut*, *blei*, *flaut*, *skreiv* og *drap* enn vi har sett i nokon av dei andre variabelgruppene.

Samla sett kan vi trygt slå fast at denne delen av studien stadfestar funn frå tidlegare studiar av bokmål (jf. Pettersen 1993; Mosling 1998; Rosén 2000; Vinde 2000; Baardsgaard 2005; Roksvold 2005; Goffeney 2013): Dei moderate variantane dominerer over dei radikale variantane i alle variabelgruppene og i den store majoriteten av einskildvariablar. Samstundes har vi sett at biletet er meir nyansert særleg når det gjeld substantiv i bunden form eintal med valfri ending *a/en*. I denne variabelgruppa, men til dels også mellom dei ortografiske variablane og mellom inkjekjønnsorda i bunden form fleirtal, har vi sett at *nokre* av dei radikale variantane dominerer temmeleg mykje over dei moderate motsvara deira, medan det for andre variablar er ei meir jamn fordeling mellom radikale og moderate variantar. Desse funna stemmer godt overeins med funna frå den tidlegare nemnde studien til Kola (2014: 57ff) av einskildvariablar. I neste kapittel skal vi sjå nærmare på det overordna forskningsspørsmålet for denne studien, nemleg om det i subkorpuset av LBK eksisterer normklynger beståande av bestemte ordformer som kan la seg identifisere som meir eller mindre dominerande subvarietetar av bokmål.

4.3 Normklynger

I ein innleiande fase av analysearbeidet gjennomførte vi først separate korrespondanseanalysar for dei språklege variablane i kvar av dei seks

tabellane ovanfor. Dei innleiande analysane peika i retning av at det innanfor tekstutvalet ville vekse fram normklynger som kunne la seg identifisere som subvarietetar av bokmål. Samstundes var det slik at dei innleiande analysane baud på problem knytte til manglande statistisk stabilitet. Derfor skal vi her gå direkte til analysen der vi jamførte eit utval av variablar frå kvar av dei seks tabellane samstundes, då denne synte seg å gje langt meir statistisk stabile resultat. Vi bestemte oss for å ta med eit breiast mogleg samansett utval av variablar frå kvar av dei seks tabellane, då likevel slik at det var om lag like stort prosentvis uttrekk frå kvar tabell, samstundes som vi freista å unngå å ta med variablar der skilnaden mellom talet på radikale og talet på moderate førekomstar var ekstrem, då dette siste kan skape ustabilitet i dei statiske analysane (McEneaney og Hardie 2012: 52).

På denne bakgrunnen gjennomførte vi ein analyse med utgangspunkt i talet på førekomstar av kvar av dei to valfrie variantane for til saman 31 variablar i kvar av dei 452 tekstane som var med i tekstutvalet. Analysen synte at χ^2 var langt over kritisk verdi. Dermed kunne vi forkaste H_0 og gå vidare med H_1 , som postulerer at avviket mellom forventa og observerte verdiar er så stor at det *er* ein statistisk samanheng mellom tekstane og formvala som ein finn i dei. Med bakgrunn i dei ovanfor nemnde funna knytte til einskildvariablar og funna i den tidlegare nemnde studien til Dyvik (2012: 218) som indikerer klyngestrukturar i eit utval avistekstar, sette vi fram ein hypotese H_2 om at det innanfor det totale tekstutvalet ville vekse fram normklynger av bestemte ordformer som kunne la seg identifisere som statistisk signifikante subvarietetar av bokmål.

Den første korrespondanseanalysen vi gjorde av dei 31 språklege variablane, synte at dei lågfrekvente, radikale ordformene *orda* og *skreiv* hadde avvikande profilar jamført med alle andre former. For å gjere den statistiske analysen meir stabil valde vi, i tråd med metodikken innanfor korrespondanseanalyse, derfor å gjere dei nemnde ordformene og dei moderate motsvara deira (*ordene* og *skrev*) om til supplementærpunkt, medan vi beheldt dei to valfrie variantane av kvar av dei 29 andre tabellane som aktive punkt. Dei supplementære variantane er passive punkt utan masse, og slike punkt yter verken bidrag til den totale variansen eller til dei einskilde aksane, og er slik ikkje i stand til å orientere aksane; dei berre plasserer seg på sine respektive plassar i høve til dei andre punkta i den allereie etablerte punktsskya (Lebart 1998: 60f; Hjellbrekke 1999: 108f; Greenacre 2007: 89f). Etter dette metodiske grepet fekk vi

Tabell 7 nedanfor, som syner kor stor del av den samla variansen av dei aktive variablane som blir fanga opp av dei åtte første aksane.

Tabell 7 Variansen til aksane, målt i eigenverdi og i prosent av den totale variasjonen

| Aksenummer | Eigenverdi | Prosent av varians | Akkumulert prosent av varians |
|------------|------------|--------------------|-------------------------------|
| 1 | 0,380 | 14,2 | 14,2 |
| 2 | 0,203 | 7,6 | 21,8 |
| 3 | 0,194 | 7,2 | 29,0 |
| 4 | 0,151 | 5,6 | 34,6 |
| 5 | 0,142 | 5,3 | 40,0 |
| 6 | 0,108 | 4,0 | 44,0 |
| 7 | 0,095 | 3,6 | 47,5 |
| 8 | 0,089 | 3,3 | 50,9 |

Vi ser at over 50 % av den totale variansen blir fanga opp av dei åtte første aksane, og at olbogekriteriet tilseier at det er nok å tolke dei tre første aksane, men før ein kan ta til å tolke resultata, treng ein som tidlegare nemnt også å finne ut kva for variantar som har vore dei viktigaste for konstruksjonen av aksane i planet. Tabell 8 nedanfor syner relativ vekt, avstand til barysenteret og bidrag til kvar av dei tre første aksane frå dei 58 aktive punkta/språklege variantane som vi la til grunn for den endelege korrespondanseanalysen.

Vi ser at akse 1 i høg grad er danna på grunnlag av opposisjonen mellom dei moderate ordformene *frem*, *boken* og *sen* og dei meir radikale formene *fram*, *boka* og *sein* som har eit samla bidrag på 71,4 %. Akse 2 blir i høg grad definert av opposisjonen mellom dei moderate ordformene *moren*, *årene*, *sen*, *foreldrene* og *snakket* og dei meir radikale formene *fram*, *sola*, *mora*, *bein* og *gata*, som har eit samla bidrag på 76,9 %. Når det gjeld akse 3, kjem dei viktigaste bidraga frå dei moderate formene *moren*, *frem*, *sen*, *foreldrene* og *årene* på den eine sida og frå dei radikale formene *sola*, *fram*, *sein*, *boka*, *mora* og *sju* på den andre, som har eit samla bidrag på 78,6 %. Sidan akse 1 er den aksan som fangar opp den klart største variansen i skya, og sidan det er ein god del overlapp mellom dei ordformene som yter størst bidrag til akse 2 og 3, vel vi å fokusere på dei samanhengane som kjem til syne i faktorplan 1-2 og i faktorplan 1-3, som blir presenterte nedanfor i Graf 1 og Graf 2. I båe dei to grafane nedanfor er det slik at dei blå punkta markerer tekstar i tekstska,

Tabell 8 Relativ vekt, avstand til barysenteret og bidrag til akse 1, 2 og 3 frå dei 58 aktive punkta/språklege variantane som vi la til grunn for den endelege korrespondanseanalysen

| Variantar | Relativ vekt | Avs. til barysenteret | Bidrag akse 1 | Bidrag akse 2 | Bidrag akse 3 |
|------------|--------------|-----------------------|---------------|---------------|---------------|
| fram | 13,578 | 1,185 | 33,9 | 11,7 | 10,6 |
| frem | 12,914 | 0,959 | 19,6 | 0,0 | 5,3 |
| gras | 0,089 | 8,743 | 0,5 | 0,7 | 2,2 |
| gress | 1,061 | 1,666 | 0,0 | 0,9 | 1,1 |
| gard | 0,096 | 3,274 | 0,0 | 0,0 | 0,2 |
| gård | 1,782 | 1,905 | 0,2 | 0,6 | 1,7 |
| bein/s | 1,861 | 1,691 | 1,9 | 4,6 | 0,0 |
| ben/s | 2,353 | 1,279 | 2,7 | 0,7 | 0,2 |
| sein | 0,838 | 3,291 | 4,3 | 2,8 | 9,4 |
| sen | 16,878 | 0,663 | 3,8 | 8,4 | 2,8 |
| skeiv | 0,062 | 5,199 | 0,1 | 0,3 | 0,0 |
| skjev | 0,571 | 1,515 | 0,1 | 0,0 | 0,1 |
| sjuk | 0,153 | 3,717 | 0,3 | 1,6 | 0,8 |
| syk | 2,731 | 1,626 | 0,1 | 0,2 | 0,0 |
| tjuv | 0,011 | 11,245 | 0,0 | 0,0 | 0,0 |
| tyv | 0,372 | 3,963 | 0,0 | 0,0 | 0,0 |
| sju | 1,682 | 2,255 | 2,0 | 0,7 | 2,6 |
| syv | 1,122 | 1,556 | 0,8 | 0,2 | 0,3 |
| verken | 2,584 | 1,147 | 0,8 | 1,6 | 0,1 |
| hverken | 1,284 | 1,856 | 1,3 | 0,2 | 0,0 |
| bygda | 0,281 | 2,951 | 0,1 | 0,1 | 0,5 |
| bygden | 0,031 | 3,244 | 0,0 | 0,0 | 0,0 |
| hytta | 0,524 | 2,723 | 0,0 | 0,8 | 0,1 |
| hytten | 0,096 | 4,410 | 0,0 | 0,1 | 0,1 |
| kona | 0,788 | 2,135 | 0,1 | 1,4 | 0,0 |
| konen | 0,200 | 2,879 | 0,3 | 0,1 | 0,0 |
| boka | 1,600 | 2,732 | 7,7 | 0,3 | 4,7 |
| boken | 2,691 | 1,590 | 2,1 | 1,0 | 0,3 |
| gata | 0,545 | 3,463 | 1,5 | 3,2 | 1,5 |
| gaten | 1,738 | 1,431 | 2,0 | 1,0 | 0,5 |
| mappa | 0,018 | 10,004 | 0,0 | 0,1 | 0,0 |
| mappen | 0,148 | 3,000 | 0,1 | 0,3 | 0,1 |
| sola | 0,766 | 4,201 | 4,9 | 5,3 | 16,0 |
| solen | 1,647 | 1,718 | 0,3 | 0,8 | 2,4 |
| mora | 0,375 | 4,242 | 1,6 | 4,8 | 3,0 |
| moren | 4,921 | 1,564 | 0,2 | 16,1 | 19,0 |
| foreldra | 0,026 | 7,783 | 0,2 | 0,3 | 0,1 |
| foreldrene | 2,814 | 2,284 | 0,7 | 5,8 | 2,6 |
| tinga | 0,016 | 6,932 | 0,1 | 0,1 | 0,1 |
| tingene | 1,470 | 1,478 | 0,1 | 0,9 | 0,2 |
| dyra | 0,101 | 6,415 | 0,5 | 0,4 | 2,1 |
| dyrene | 0,827 | 2,325 | 0,3 | 0,0 | 0,0 |
| ben | 1,215 | 1,420 | 1,4 | 0,8 | 0,3 |
| benene | 0,022 | 3,734 | 0,0 | 0,0 | 0,0 |
| fjella | 0,015 | 11,288 | 0,1 | 0,0 | 0,3 |
| fjellene | 0,864 | 2,088 | 0,0 | 0,0 | 0,1 |

| | | | | | |
|---------|-------|-------|-----|------|-----|
| åra | 0,244 | 4,901 | 1,4 | 0,8 | 2,1 |
| årene | 6,098 | 1,150 | 0,2 | 11,2 | 2,6 |
| drap | 0,048 | 3,350 | 0,0 | 0,0 | 0,0 |
| drepte | 1,225 | 2,069 | 0,0 | 0,0 | 0,8 |
| angra | 0,008 | 8,304 | 0,0 | 0,1 | 0,0 |
| angret | 0,226 | 1,968 | 0,1 | 0,1 | 0,2 |
| jobba | 0,064 | 4,417 | 0,2 | 0,6 | 0,2 |
| jobbet | 1,054 | 1,472 | 0,1 | 0,6 | 0,2 |
| mangla | 0,031 | 5,630 | 0,1 | 0,1 | 0,1 |
| manglet | 0,797 | 1,346 | 0,2 | 0,0 | 0,0 |
| snakka | 0,127 | 4,444 | 0,3 | 1,5 | 1,1 |
| snakket | 4,317 | 1,110 | 0,8 | 5,8 | 1,1 |

medan raude og fylte trekantar markerer aktive språklege variantar, og raude og ufylte trekantar markerer passive språklege variantar, i ordsksya.

Hovudbiletet som teiknar seg, stadfestar hypotesen H_2 om at det innanfor det samla tekstutvalet som ligg til grunn for studien, ville vekse fram normklynger av bestemte språklege variantar som kunne la seg identifisere som subvarietetar av bokmål. For det første syner samleanalysane i Graf 1 og Graf 2 at dei aller fleste moderate ordformene og eit mindre tal radikale ordformer hamnar nær barysenteret med relativt liten innbyrdes avstand, noko som inneber at dei er høgfrekvente og har ein tendens til å opptre i lag på tvers av mange tekstar. Ifølgje metodikken innanfor korrespondanseanalyse gjev dette grunnlag for å slå fast eksistensen av ei dominerande og statistisk signifikant normklynge i tekstutvalet. Konsentrasjonselipsane i dei to grafane er berekna og innteikna av den statiske korrespondanseanalysen, og om ein ser på dei to ellipsane som ligg nærmast barysenteret i Graf 1 og Graf 2, avdekkjer dei eksistensen av ei normklynge som hovudsakleg består av moderate former som til dømes *gress*, *syk*, *fjellene*, *tingene*, *angret* og *snakket*, men også av nokre få meir radikale former som *bygda* og *hytta*.

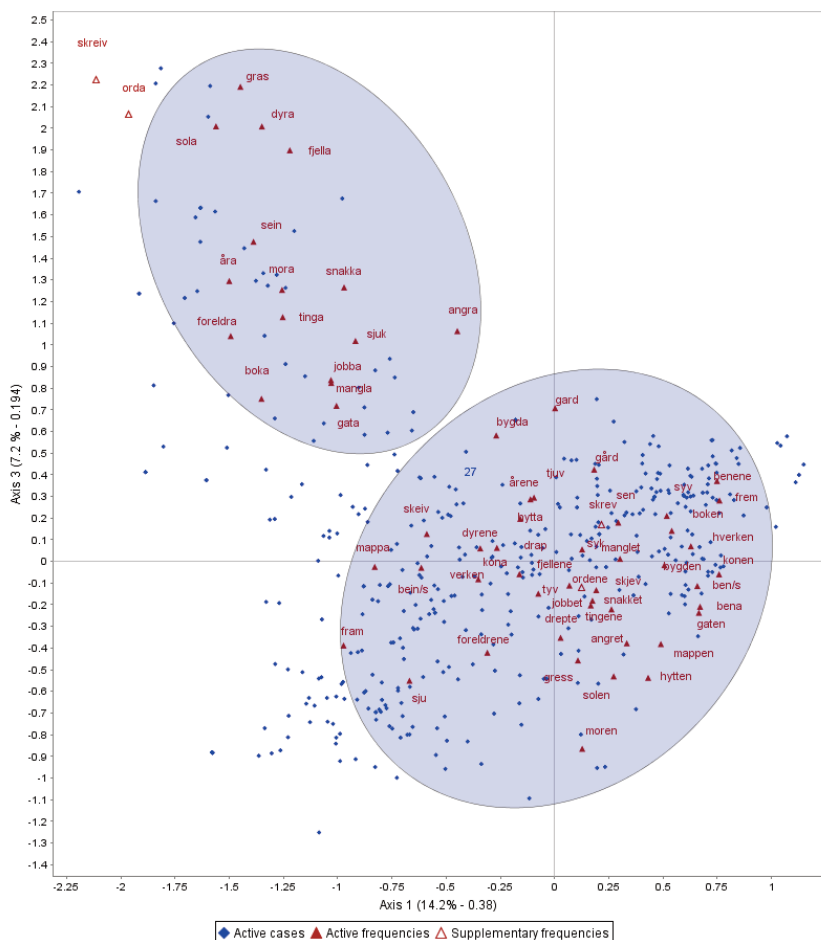
For det andre ser vi at dei fleste radikale formene hamnar innanfor dei to ellipsane som hamnar lengre vekk frå barysenteret. Dette indikerer at det mellom eit mindre tal av tekstane i tekstutvalet eksisterer ei normklynge bestående av former som *gras*, *sjuk*, *gata*, *fjella*, *jobba*, *tinga*, *angra* og *snakka*. Likevel er den innbyrdes avstanden mellom iallfall ein del av desse formene så pass stor at ein ikkje utan vidare kan slå fast at det eksisterer ei statistisk signifikant normklynge av alle desse radikale formene. Særleg har formene *skreiv* og *orda* så avvikande profilar i høve til resten av dei radikale formene at det er grunn til å stille spørsmål ved om dei tilhøyrer ei eventuell normklynge av radikale former.

Graf 1 Radikale og moderate språklege variantar. Faktorplan 1-2.



Vi legg merke til at formene *verken* og *hverken* hamnar relativt nær kvarandre i rommet og innanfor det vi har slått fast er ei signifikant normklynge som hovudsakleg består av moderate former, men altså også nokre radikale former. Dette let seg dels forklare ved det faktum at det er nokså jamn fordeling mellom dei to formene i det samla tekstutvalet – jf. Tabell 1 ovanfor, som syner at 66,8 % av førekomstane av denne variabelen er på forma *verken*, medan 33,2 % er på forma *hverken*, og dels

Graf 2 Radikale og moderate språklege variantar. Faktorplan 1-3.



ved at den opphavlege resultatmatrisa avdekkjer at så pass mykje som 11,0 % av dei i alt 397 skribentane som i det heile har brukt det aktuelle ordet, vekslar mellom å bruke formene *verken* og *hverken* i ulike tekstar, medan 42,8 % held seg konsekvent til forma *hverken* og 46,0 % held seg konsekvent til forma *verken*.

Noko meir overraskande er det at også formene *gard* og *gård* og *drap* og *dreppte* hamnar relativt nær kvarandre i rommet og innanfor den normklynga som hovudsakleg består av moderate former. Dei radikale for-

mene *gard* og *drap* er nemleg svært lite brukte – jf. Tabell 1 ovanfor, som syner at *gard* har ein relativ frekvens på berre 5,1 %, medan Tabell 6 ovanfor syner at *drap* har ein relativ frekvens på berre 3,8 % i det samla tekstutvalet. At desse formene hamnar nær kvarandre i rommet, let seg likevel forklare dels ved at ein relativ stor del av skribentane vekslar mellom *gard* og *gård* og mellom *drap* og *drepte*, og dels ved at Tabell 8 ovanfor syner at desse formene i svært liten grad påverkar konstruksjonen av akse 1, 2 og 3 i dei to grafane.

Analysen viser også at nokre former plasserer seg imellom dei ulike klyngene som er prega av høvesvis moderate og radikale former. Det understrekar at eventuelle normklynger ikkje er kategoriske storleikar, men at det heller er tale om klynger av prototypisk karakter. At mange skribentar brukar mange av dei same moderate formene, til dømes, tyder ikkje at alle desse brukarane av moderat bokmål brukar alle dei same moderate formene. Nokre av dei aktuelle skribentane kan bruke den og den radikale forma og likevel halde seg innanfor det som er rimeleg å rekne for å vere ei moderat normklynge. Døme på former som plasserer seg i grenselandet mellom normklynga bestående av hovudsakleg moderate former og den moglege normklynga bestående av radikale former i Graf 1, er *sju*, *kona* og *tjuv*. Det er slik statistisk sannsynleg at ein god del av skribentane som elles brukar moderate former, også brukar desse tre meir radikale formene, noko ein gjennomgang av den opphavlege resultatmatrisa også indikerer.

Det er grunn til å presisere at desse korrespondanseanalysane er pålitelege og robuste ut frå statistiske kriterium. Som vi ser, er det nemleg stor spreining av punkta i dei to skyene både i Graf 1 og i Graf 2, noko som igjen impliserer stabilitet, dvs. at det er høg statistisk samanheng mellom verdiane i materialet (Hjellbrekke 1999: 37). Den statistiske stabiliteten kjem også til uttrykk gjennom Tabell 8 ovanfor, som syner at dei ulike aksane i dette tilfellet blir danna på grunnlag av mange ordformer. På denne måten kan ein seie at funna i denne studien klart styrker indikasjonane på eksistensen av ei moderat normklynge innanfor bokmål som Dyvik (2012: 218) fann i den tidlegare nemnde studien av eit utval avistekstar. Alt i alt lyt dette bli karakterisert som statistisk sikre og til dels slående funn som etter vår vurdering bør få implikasjonar for det framtidige normeringsarbeidet for bokmål, noko vi vil kome attende til i neste kapittel.

5 Oppsummering

Noreg har lenge ført ein særeigen og sjeldsynt normeringspolitikk prega av tradisjonsprinsippet versus tilnæringsprinsippet på den eine sida og fastleiksprinsippet versus folkemålsprinsippet og valfridomsprinsippet på den andre. Resultatet var at vi ved inngangen til 2000-talet hadde veldig stor innbyrdes valfridom mellom hovudformer og sideformer innanfor baa dei to offisielle norske skriftspråka. Men i åra etter millenniumskiftet kom det ei dreiging i den offisielle språkpolitikken, i og med at Stortinget (St.meld. nr. 9 2001-2002: 25) vedtok at tilnæringslinja ikkje lenger skulle vere gjeldande politikk. I staden skal mål- dyrkingsarbeidet no vere «knytt til spørsmålet om korleis bokmål og nynorsk kvar for seg skal kunne utvikle seg best mogleg.» Med dette som utgangspunkt blei det vedteke ei ny rettskriving for bokmål i 2005 som var danna på sjølvstendig grunnlag, uavhengig av nynorsk, samstundes som det tidlegare skiljet mellom hovudformer og sideformer vart oppheva. Dessutan blei ei rekkje radikale former som tidlegare var tillatne, tekne ut av rettskrivinga, samstundes som ei lang rekkje tradisjonelle og konservative former som ikkje hadde vore tillatne sidan før 1938, blei tekne inn att i rettskrivinga. Samstundes er det framleis valfridom mellom ulike ordformer innanfor svært mange språklege variablar i gjeldande bokmålsnormal.

Hausten 2021 vedtok styret i Språkrådet (2021) nye *Retningslinjer for normering av bokmål og nynorsk*. I desse retningslinjene heiter det at «[d]et vesentlege grunnlaget for normeringa av bokmål er språket slik det ser ut i bokmålstekstar», og at normeringsgrunnlaget i første rekkje skal vere «tekstar som har vore gjennom ein redigeringsprosess» (Språkrådet 2021: 12). Med det til siktemål å styrke kunnskapsgrunnlaget for den vidare normeringa av bokmål har vi derfor gjennomført ein studie der vi har kartlagt bruken av ulike valfrie ordformer mellom 150 språklege variablar i eit breitt samansett utval av 452 redigerte og publiserte tekstar som ligg i Leksikografisk bokmålskorpus (LBK). Det overordna forskingsspørsmålet for studien var om det i tekstutvalet som ligg til grunn for studien, eksisterer normklynger beståande av bestemte ordformer som kunne la seg identifisere som meir eller mindre dominerande subvarietetar av bokmål.

Før vi søkte svar på dette forskingsspørsmålet, gjennomførte vi ei kartlegging av tilhøvet mellom bruken av moderate og radikale former

innanfor einskildvariablar som stadfesta funn frå tidlegare studiar av bokmål (jf. Pettersen 1993; Mosling 1998; Rosén 2000; Vinde 2000; Baardsgaard 2005; Roksvold 2005; Goffeney 2013): Dei moderate variantane dominerer over dei radikale variantane i alle variabelgruppene og i den store majoriteten av einskildvariablar. Samstundes syner denne delen av studien at biletet er meir nyansert, særleg når det gjeld substantiv i bunden form eintal med valfri ending *a/en*. I denne variabelgruppa, men til dels også mellom dei ortografiske variablane og mellom inkjekjønnsorda i bunden form fleirtal, er det nemleg slik at *nokre* av dei radikale variantane dominerer temmeleg mykje over dei moderate motsvara deira, medan det for andre variablar er ei meir jamn fordeling mellom radikale og moderate variantar. Desse funna stemmer godt overeins med funna i ein studie av Kola (2014: 57ff) av einskildvariablar.

Studien av normklynger styrker indikasjonane på eksistensen av ei moderat normklynge innanfor bokmål som Dyvik (2012: 218) fann i ein mindre studie av eit utval avistekstar. Korrespondanseanalysane i Graf 1 og Graf 2 i resultatkapittelet ovanfor syner nemleg at dei aller fleste moderate ordformene og eit mindre tal radikale ordformer hamnar nær barysenteret med relativt liten innbyrdes avstand, noko som gjev grunnlag for å slå fast eksistensen av ei dominerande og statistisk signifikant normklynge i tekstutvalet som hovudsakleg består av moderate former som til dømes *gress, syk, fjellene, tingene, angret* og *snakket*, men også av nokre få meir radikale former som *bygda* og *hytta*.

Når ein samanheld funna frå kartlegginga av tilhøvet mellom bruken av moderate og radikale former innanfor einskildvariablar og kartlegginga av normklynger, ser vi at den operative norma som nedfeller seg i det store og samansette tekstutvalet som ligg til grunn for denne studien, består av ei normklynge som i hovudsak består av moderate former, men også nokre radikale former. Når vi samstundes veit at den offisielle norske språkpolitikken tilseier at ein skal basere normeringa av bokmål nettopp på slike redigerte og publiserte tekstar som ligg til grunn for denne studien, talar funna i denne studien for at ein ved framtidig normering går lenger i retning av det som her er omtala som moderat bokmål.

Litteratur

- Benzécri, Jean-Paul. 1969. Statistical analysis as a tool to make patterns emerge from data. I: *Methodologies of pattern recognition*, red. Shinji Watanabe, 35–74. New York: Academic Press.
- Benzécri, Jean-Paul. 1973. *L'Analyse des Données. Vol. 1: Taxionomie. Vol. 2: Analyse des Correspondances*. Paris: Dunod.
- Baardsgaard, Øystein. 2005. *En kjernenormal i bokmålet?: en undersøkelse av språket i åtte skjønnlitterære verk*. Hovedoppgave, Universitetet i Oslo.
- Dyvik, Helge. 2009. Normering i den postsammnorske æra. I: *Språknormering – i tide og utide?*, red. Helge Omdal og Rune Røsstad, 109–121. Oslo: Novus forlag.
- Dyvik, Helge. 2012. Norm clusters in written Norwegian. I: *Exploring Newspaper Language. Using the web to create and investigate a large corpus of modern Norwegian*, red. Gisle Andersen, 193–219. Amsterdam: John Benjamins Publishing Company.
- Fjeld, Ruth Vatvedt. 2015. Språklige varieteter eller språklige rarieteter? Om bruk av valgfrie former i norsk bokmål. *LexicoNordica* 22, 35–56.
- Fjeld, Ruth Vatvedt, Anders Nøklestad og Kristin Hagen. 2020. Leksikografisk bokmålskorpus. Bakgrunn og bruk. I: *Leksikografi og korpus. En hyllest til Ruth Vatvedt Fjeld*, red. Janne Bondi Johannessen og Kristin Hagen, *Oslo Studies in Language* 11 (1), 101–124.
- Goffeney, Claude. 2013. "Eg får ikkje til nynorsk, og e lika ikkje nynorsk, og e lika ikkje å les det, og før du sei det, e veit e prata det": Forholdet mellom tale og skrift på Ytre Helgeland: skriftspråkleg praksis, talemål og haldningar. Masteroppgåve, Høgskulen i Volda.
- Greenacre, Michael J. 2007. *Correspondence Analysis in Practice*, Second Edition. London and New York: Chapman & Hall.
- Guttu, Tor. 2005. Rettskrivingsendringene i 2005. *Språknytt* hf. 3, 1–4.
- Helset, Stig J. 2017. *Norm og røyndom – ein statistisk studie av operative normer i det nynorske skriftspråket*. Doktoravhandling, Universitetet i Bergen.
- Hjellbrekke, Johs. 1999. *Innføring i korrespondanseanalyse*. Bergen: Fagbokforlaget.

- Kola, Kjersti Wictorsen. 2014. *Bokmålsbruk – hvorledes/hvordan/åssen og hvorfor? Om bruken av ortografiske og morfologiske varianter i bokmålsnormalen*. Masteroppgave, Universitetet i Oslo.
- Kulturdepartementet. 2001–2002. Målbruk i offentlig teneste. St. meld. nr. 9 (2001–2002). Oslo: Kulturdepartementet. Henta 2.2.22 <https://www.regjeringen.no/nb/dokumenter/stmeld-nr-9-2001-2002-/id195336/>
- Kulturdepartementet. 2019–2020. Lov om språk (språkløva). Prop. 108L. (2019–2020). Oslo: Kulturdepartementet. Henta 2.2.22 <https://www.regjeringen.no/no/dokumenter/prop.-108-l-20192020/id2701451/>
- Lebart, Ludovic 1998. *Exploring Textual Data*. Dordrecht: Kluwer Academic Publishers.
- Le Roux, Brigitte og Henry Rouanet. 2010. *Multiple Correspondence Analysis*. London: SAGE.
- McEnery, Tony og Andrew Hardie. 2012. *Corpus Linguistics. Method, Theory and Practice*. Cambridge: Cambridge University Press.
- Mosling, Atle. 1998. *Språknormering i sakprosaen*. Hovedoppgave, Universitetet i Oslo.
- Pettersen, Egil. 1993. *Språknormering og forfatterne: Ortografi og morfembruk hos ti bokmålsforfattere for hvert av årene 1937, 1957 og 1977*. Nordisk institutt, Universitetet i Bergen.
- Rauset, Margunn. 2022. Merkelappar brukte på bokmål og nynorsk med subvarietetar. I: *Han Johan og dei. Veneskrift til Johan Myking ved syttiårsleitet*, red. Randi Neteland og Torodd Kinn, *BeLLS* 12 (1), 207–232.
- Roksvold, Thore. 2005. *Var avisspråket bedre før? En metodetriangulert studie av normendringer i norsk avisspråk på 1900-tallet*. Doktoravhandling, Universitetet i Oslo.
- Rosén, Victoria. 2000. Er norsk et naturlig språk?. I: *Menneske, språk og fellesskap. Festskrift til Kirsti Koch Christensen på 60-årsdagen 1. desember 2000*, red. Øivin Andersen, Kjersti Fløttum & Torodd Kinn, 158–173. Oslo: Novus.
- Språkrådet. 2021. *Retningslinjer for normering av bokmål og nynorsk*. Henta 2.2.22 <https://www.sprakradet.no/Spraka-vare/Norsk/normering/retningslinjer-for-normering/>
- Tekstlaboratoriet. 2022. *Leksikografisk bokmålskorpus*. <https://www.hf.uio.no/iln/tjenester/kunnskap/sprak/korpus/skriftsprakskorpus/lbk/>
- Vedtekter for Språkrådet. Henta 2.2.22 <https://www.sprakradet.no/Vi-og-vart/Om-oss/Vedtekter-for-Sprakradet/>

- Vinde, Eva Cathrine Kolsrud. 2000. *Syv eller sju? – og mye mer: Språknormer og språkpraksis i norsk presse*. Hovedoppgave, Universitetet i Oslo.
- Vinje, Finn-Erik. 2004. *Språksituasjonen i Norge i historisk perspektiv*. Oslo: s.n.

Abstract

This article examines how permitted variability within official Bokmål is reflected in the operative norms in a large and broad selection of edited Bokmål texts. In the autumn of 2021, the board of the Norwegian Language Council adopted new guidelines for the standardization of Bokmål and Nynorsk. In these guidelines it is stated that «[t]he essential basis for the standardization of Bokmål is as Bokmål appears in Bokmål texts», and that the basis for standardization should primarily be «texts that have been through an editing process». With the aim of strengthening the empirical basis for the further standardization of Bokmål, we have carried out a study in which we have mapped the use of various optional word forms between 150 linguistic variables in a broad selection of 452 edited and published texts contained in the *Leksikografisk bokmålskorpus*. In addition to recording and comparing frequencies of use for different optional forms within individual variables, we have used correspondence analysis to map norm clusters of different subvarieties of Bokmål. The main conclusion is that the moderate forms dominate over the radical forms for the vast majority of the individual variables included in the study, at the same time as the radical forms dominate over the moderate forms for a few of the individual variables, and not least that between the texts in the sample included in the study, there is a norm cluster of what can be characterized as moderate Bokmål.

Stig J. Helset
Høgskulen i Volda
Institutt for språk og litteratur
Postboks 500, NO-6101 Volda
stig.jarle.helset@hivolda.no