



## Innleiing

Ved eit fagleg seminar ved UiB hausten 2020 vart det markert at Språksamlingane no er på plass i Bergen. Målet med seminaret var å visa kva ressurs Språksamlingane er, både i språkforskning og i anna forskning. Det er også målet med dette temanummeret av *Maal og Minne* – det første i den lange historia til dette tidsskriftet.

Innleiinga til temanummeret er skriven av Peder Gammeltoft, leiar av Språksamlingane og Randi Neteland, leiar av Fagrådet for Språksamlingane. Dei presenterer dei digitale og fysiske ressursane knytt til dei fem ulike delane av samlingane: leksikografi, stadnamn, terminologi, målføre og norrønt, og skisserer nokre av dei moglegheitene og utfordringane som ligg i tida som kjem. Gammeltoft og Neteland presenterer også innhaldet i dei fem andre artiklane, like mangfaldig som samlingane dei tek utgangspunkt i. Vi viser lesaren vidare til innleiinga deira – og til dei meir spesifikke bidraga.

God lesnad!

Hans-Olav Enger og Ingvil Brügger Budal





# Språksamlingane – ein mangfaldig forskingsressurs

Peder Gammeltoft og Randi Neteland

## Innleiing

Språksamlingane er i dag etablert som ein nasjonal ressurs for utforskinga av norsk språk. Vi kan grovt dele innhaldet i samlingane i fem hovudområde: Leksikografi, stadnamn, terminologi, målføre og norrønt. I kvart av desse delområda finn vi empiriske data om norsk språk som er samla inn over ein lang tidsperiode av ulike forskarar og forskingsprosjekt med røter i ulike forskingstradisjonar. Språksamlingane er i utgangspunktet ei samling av ulike og mangfaldige arkiv som hovudsakleg blei oppretta i første halvdel av 1900-talet (sjå nedanfor). Desse har over tid blitt utbygd og supplert – og seinare ganske omfattande digitalisert t.d. gjennom *Dokumentasjonsprosjektet* (1992–1997) og seinare *Eining for digital dokumentasjon* (EDD) ved Universitetet i Oslo (UiO). I 2016 blei dei fleste av språksamlingane ved UiO overført til Universitetet i Bergen (UiB) i ein massiv flytteprosess og samla under nemninga *Språksamlingane* (sjå meir om dette i Mykings artikkel i dette nummeret).<sup>1</sup> Då stod det klart at dei fysiske samlingane aleine utgjorde 1,8 km arkivhyllometer og dei digitaliserte delane 29 terabyte. Altså var det ei monumental oppgåve UiB tok på seg.

Språksamlingane er med andre ord ein samansett og mangfaldig forskingsressurs som det kan vere vanskeleg å få oversikt over. I tillegg til å vere ein forskingsressurs er Språksamlingane også brukt i forvaltninga av stadnamn, og Språksamlingane driftar nettordbøkene *Bokmålsordboka*

1. I 2014 vart det utarbeidd ei oversikt over språksamlingane ved UiO i rapporten «Den forskings- og undervisningsmessige verdien av eit utval samlingsressursar ved ILN» (Ebeling m.fl. 2014). **Vedlegga** til denne rapporten gir detaljert informasjon om innhaldet i dei språklege samlingane som den gong var ved UiO. Mesteparten av desse blei overførte til UiB.

og *Nynorskordboka* som er godt kjend og brukt av ålmenta. I dette tema-nummeret av *Maal og Minne* får vi døme på korleis ulike delar av Språksamlingane kan takast i bruk i moderne språkforsking. Denne korte innleiinga har som føremål å gi ei enkel oversikt over dei ulike delane av samlingane og kva desse delsamlingane inneheld av språkleg materiale.

## Digitale og fysiske ressursar

Språksamlingane består både av digitaliserte og fysiske samlingar, men det er ikkje fullstendig overlapping mellom dei digitale samlingane og dei fysiske arkiva. Det finst meir i digital form enn det som ligg i arkiva, og likeins inneheld arkiva ein god del meir enn det som er digitalisert. Samlingane inneheld alt det språkmaterialet som blei overført frå UiO (sjå Ebeling m.fl. 2014), og har blitt tilført meir materiale i seinare tid.

Det digitaliserte materialet som blei overført frå samlingane i Oslo, bestod i hovudsak av setelsamlingar og kartsamlingar, og i mindre grad av innskanna heile verk, slik som Målføresynopsisen. I tillegg til at dei er skanna, finst setelsamlingane òg i digitalisert tilstand, det vil seie at dei skanna og avfotograferte setlane er blitt transkribert og innhaldssystematisert. Det same er tilfellet for den store mengda sentrale ordbøker og dialektordlister som inngår i metaordboka til nettutgåva av *Norsk Ordbok*. Av dei norrøne delane av samlingane er ein stor del innlemma i *Menota – Medieval Nordic Text Archive* ([www.menota.org](http://www.menota.org)), liksom digitaliserte utgåver av Riksarkivets *Diplomatarium Norvegicum* og Oluf Ryghs *Norske Gaardnavne*.

Då dei fysiske arkiva opphavleg bestod av ei rekke uavhengige samlingar, er dei aldri blitt registrert og ordna etter felles og einsarta prinsipp. Dette er ei av hovudoppgåvene for Språksamlingane – ei ganske undereksponert oppgåve som er heilt essensiell for å kunne vurdere kva som kan og bør digitaliserast i framtida. Etter kvart som dei individuelle samlingane i det fysiske arkivet blir ferdig ordna og registrert, blir dei publisert på *Arkivportalen*.<sup>2</sup> Vi vil presisere at manglande digitalisering ikkje

2. Med tida vil det vere mogleg å sjå Språksamlinganes arkivmateriale inndelt etter leksikografi, namnegransking, målføre og norrønt på *Arkivportalen* – i skrivande stund er berre eit fåtal av samlingane registrert der. Sjå <https://www.arkivportalen.no/search?repository=SPR> for ferdigordna materiale.

er noko hinder for å bruke materialet i forskning – det er berre å ta kontakt med Språksamlingane og be om å få sjå materialet. Til dømes har Benthe Kolberg Jansson og Karine Stjernholm frå Høgskulen i Østfold brukt arkivmaterialet etter talemålsforskarer Ingeborg Hoff i forskinga si på østfoldmålet (Jansson 2018).

I skrivande stund er òg arkiva og samlingane ved Humanistisk fakultet på UiB på veg å bli overført til Spesialsamlingane ved Universitetsbiblioteket. I samband med overføringa kjem dei språklege samlingane, altså Stadnamnarkivet og Målføresamlinga, til å inngå i Språksamlingane. Vi får såleis omsider samla alle språkressursane ved UiB i Språksamlingane (sjå Mykings artikkel i dette nummeret, sjå også oversikt i *Arkivportalen* (fotnote 2) og oversiktsrapporten frå UiO (Ebeling m.fl. 2014)).

## Leksikografi

Leksikografi og namnegransking har heilt frå starten av vore ei prioritert arbeidsoppgåve for Språksamlingane (jf. tildelingsbrevet frå Kunnskapsdepartementet 2015: 8). Ordning og registrering av dei fysiske samlingane har difor fokusert på desse områda.<sup>3</sup> Det leksikografiske arbeidet blir først og fremst utført gjennom dei to store prosjekta *Revisjon av Bokmålsordboka og Nynorskordboka (Revisjonsprosjektet)* (2018–2023), som har som oppgåve å revidere og oppdatere standardordbøkene *Nynorskordboka* og *Bokmålsordboka*, og *Norsk Ordbok a–h (NO–AH)* (2019–2029), som har som oppgåve å revidere og digitalisere ferdig *Norsk Ordbok – ordbok over det norske folkemålet og det nynorske skriftmålet*. Som eit resultat av desse prosjekta blir nettordbøkene som blir brukt av ålmenta, modernisert og oppdatert, og det same skjer med dei forskingsressursane som leksikografane treng. Til dømes er ei rekke korpus som inneheld norsk språk, samla i korpusgrensesnittet *Corpuscle* i CLARINO (Meurer 2012; Lyse 2021). Dette inkluderer til dømes *Nynorskorpuset*, *Aviskorpuset* for bokmål og nynorsk, og eit stort korpus over frie tekstar i Nasjonalbiblioteket. Desse korpusa er heilt sentrale i det daglege arbeidet i dei leksikografiske prosjekta, til dømes for å dokumentere bruk,

3. Saman med målføre, då målføresamlingane i realiteten er så tett knytt opp mot det leksikografiske arbeidet at det pragmatisk sett ikkje gir mening å sjå målføresamlingane som åtskilte frå leksikografien, jf. vedlegg 3 og 4 i oversikta frå UiO 2014.

finne tydingar og illustrerande bruksdøme. I artikkelen til Gunn Inger Lyse, Margunn Rauset og Helge Dyvik i dette nummeret kan du lese om korleis dei brukar korpusa *Korpuskel-leks* og *NorGramBank* til å definere og studere kollokasjonar. I Leiv Inge Aa sin artikkel i dette nummeret tek han utgangspunkt i *Setelarkivet* og *Nynorskkorpuset* for å kartlegge den teoretiske statusen til infinitivsmarket *å* og drøfte kva leksikografiske oppføringar om *å* bør innehalde. Dei leksikografiske korpusa blir brukt i mange slags forskning på norsk skriftspråk. Nokre andre døme er Andersen (2012, red.) der ein kan lese om *Aviskorpuset* og ulike casestudiar som brukar korpuset som ressurs, og Lapponi, Søyland, Vellidal og Oepen (2018) for ein presentasjon av *Talk of Norway*-korpuset med stortingsforhandlingar og døme på korleis korpuset kan brukast i forskning. Alle desse leksikografiske forskingsressursane er tilgjengelege via nettstaden <https://clarino.uib.no/korpuskel/corpus-list>.<sup>4</sup>

## Stadnamn

Stadnamn er eit anna prioritert område for Språksamlingane. Samtidig med ordning og registrering av det fysiske arkivet som blei overført frå Oslo (jf. Ebeling m.fl. 2014: 7–10 og vedlegg 2), er det òg gjort ei omfattande digitalisering av særleg dei topografisk inndelte stadnamnsamlingane. I 2018 overtok Språksamlingane dessutan Noregs Teknisk-Naturvitskaplege Universitet (NTNU) sitt stadnamnarkiv og innlemma det i dei fysiske samlingane, og i det små er det starta ei digitalisering av desse samlingane. I 2019 tok også Språksamlingane på seg å skanne Universitetet i Stavanger (UiS) sitt stadnamnarkiv før det skulle overførast til eit felleskommunalt arkiv, og kort tid etter blei det inngått eit samarbeid med lokale krefter i Rogaland om digitalisering av samlingane. Digitaliseringsprosessen kan følgjast på <https://lokalnamn.no/>, og i skrivande stund er allereie om lag 130 000 stadnamn digitalisert.

Nye digitale kjelder til stadnamnnormering og stadnamngransking er også blitt innlemma og oppgradert til å kunne inngå i dei nye stadnamnløysingane som er utvikla av Språksamlingane. Desse nye ressursane er mellom anna stadnamnportalen *Norske stadnamn | Norske stedsnavn* (<https://toponymi.spraksamlingane.no>). Dette er ei nettløysing som kan

4. Til nokre fåe av korpusa (til dømes dei med talemålsdata) må ein søke om tilgang.

brukast både av forvaltninga, språkforskarar og alle andre interesserte. Ein kan til dømes søke opp eit stadnamn eller ein namnetype og få sjå utbreiinga av dette på eit kart, og også få tilgang til å lese kva det står i notidige kjelder (t.d. Kartverket) og namnegranskingskjelder (t.d. Oluf Ryghs *Norske Gaardnavne*) om desse stadnamna. Sjå nærare skildring av denne ressursen i Gammeltoft (2022), sjå også Gammeltoft (under publisering) og Vasshus (under publisering) for andre døme på forskning som tek utgangspunkt i stadnamnressursane i Språksamlingane.

## Terminologi

Terminologi er eit delområde i Språksamlingane som er forholdsvis nytt. Denne delen følgde ikkje med samlingane som blei overført frå Universitetet i Oslo, men stammar frå *Termportalen* ved Noregs Handelshøgskule (NHH). I 2018 blei drifta av terminologiresursen Termportalen overført frå NHH til Språksamlingane, og terminologi og fagspråk blei dermed gjort til den tredje prioriteten i samlingane. Dette er ei satsing som ber frukter med ei fast rammeløyving over statsbudsjettet frå 2021 til utvikling av Termportalen som nasjonal terminologisk infrastruktur. Med Termportalen skal det utviklast ein infrastruktur for terminologi og fagspråk der allereie etablerte og nyskapte terminologiske ressursar blir integrert via éi felles søkerute (<https://term.uib.no>). Det er mogleg å søkje både i alle ressursane på ein gong og i spesifikke termbasar (jf. Andersen og Gammeltoft (under publisering); Andersen og Kristiansen (2015); sjå òg til dømes Kristiansen (2014) for døme på terminologisk forskning som ligg til grunn for *Termportalen*). Termportalen har òg ein redigeringsmodul oppbygd etter websemantiske prinsipp, der fagfolk kan skrive inn, redigere og kople terminologi frå sitt fagfelt. Målet er å samle flest mogleg fritt tilgjengelege ressursar i denne portalen og at den skal bli aktivt brukt av fagfolk, studentar, journalistar og ålmenta. Universitets- og høgskulelova seier at høgare utdanning i Noreg skal utvikle og ivareta norsk terminologi. Gjennom Termportalen tek UiB og Språksamlingane sin del av dette nasjonale ansvaret.

## Målføre

Språksamlingane har også eit stort dialektologisk materiale som først og fremst kjem frå Norsk Målførearkiv i Oslo (jf. Ebeling m.fl. 2014: 12–15 og vedlegg 4).<sup>5</sup> Dei viktigaste forskingsressursane som er her, er Målføresynopsisen og Norsk dialektatlas og bakgrunnsdataa for dette prosjektet.<sup>6</sup> *Norsk dialektatlas* var eit prosjekt som blei initiert i 1951, og planen var då å lage eit dialektatlas etter mønster frå andre lands dialektatlas (Hoff 1960). Av grunnar som er ukjende for oss, blei aldri dette atlaset publisert, men mange av karta er ferdig teikna. Det finst 961 dialektkart med isoglossar som er innskanna og digitalt tilgjengelege, og dei fysiske karta ligg i arkivet. Mesteparten av dialektatlasets blei laga på bakgrunn av dialektmateriale som var samla inn som utfylling av spørjelisten *Storms kortere ordliste* (1884). Denne ordlista på over 1500 ord gir oversikt over fonologi og morfologi i ein einskild dialekt, samt nokre syntaktiske konstruksjonar. Slike ordlister blei samla inn over heile landet, og svara på desse spørjelistene finst i Språksamlingane sine arkiv. Dette la også grunnlaget for *Målføresynopsisen*. Dette er ei systematisk oppstilling av orda i ordlista, sortert geografisk, slik at ein til dømes kan samanlikne nabodialektar.<sup>7</sup> Målføresynopsisen skulle i utgangspunktet vere eit grunnlag for å teikne karta i Norsk dialektatlas (Hoff 1960), men har no blitt ein sjølvstendig forskingsressurs og eit oppslagsverk som mange talemålsforskarar nyttar seg av, til dømes som referanse til korleis den tradisjonelle dialekten var på ein bestemt stad (jf. Wetås 2014). Sjå Netelands artikkel i dette nummeret for ei nærare skildring av Målføresynopsisen som forskingsressurs og døme på korleis synopsisen kan brukast til å stille saman dialektopplysningar for mange stader.

5. I skrivande stund er også materialet frå Målføresamlinga ved Universitetet i Bergen i ferd med å innlemmast i desse samlingane.
6. Målføredelen av Språksamlingane inneheld altså ikkje lydopptak. Dette er fordi opptaka frå Målførearkivet i Oslo og dei fleste andre dialektologiske opptaka frå målføresamlingane rundt omkring i landet blei digitalisert gjennom prosjektet *Language Infrastructure made Accessible (LIA)* og gjort tilgjengelege frå nettstaden <http://www.tekstlab.uio.no/LIA/>. Mange av desse opptaka er transkribert og ligg i ei søkbart korpus, og dei som ikkje vart transkribert, er digitalt tilgjengeleg via fildepotet.
7. Ifølge Ingeborg Hoff er mesteparten av materialet i Målføresynopsisen samla inn mellom 1945 og 1960, og informantane er midaldra eller eldre (Hoff 1960: 597). Materialet skal difor vere ganske einsarta og samanliknbart på tvers av stadene.



Målføresynopsisen er også brukt som ressurs i dei leksikografiske prosjekta ved UiB. Dette har skapt behov for ein søkbar database, der ein kan søke enkelt og sortere på ord, grammatiske kategoriar eller stad. Det blei avsett midlar til dette gjennom prosjektet Norsk Ordbok A–H, og arbeidet med å «nydigitalisere» målføresynopsisen er no i gang. Vi håpar at ein ny og brukarvennleg søkeportal kan vere tilgjengeleg for alle talemålsforskarar om ikkje lenge (på <https://synopsis.uib.no>). Førebels er det dei innskanna bileta av synopsisen som er digitalt tilgjengelege.

Språksamlingane har også eit stort arkiv med oversikt over ordtilfang i dei norske dialektane (dei såkalla A-listene), samt mindre samlingar om ordlaging (D-listene), fonologi (E- og F-listene) og morfologi (G-listene). Desse blei samla inn gjennom eit landsdekkande medarbeidarnett som fekk tilsendt spørsmål frå Norsk målførearkiv. Listene er lite utnytta i forskning, og er ikkje digitalt tilgjengelege førebels, men tilgjengelege frå arkivet.

Det finst også ein del forskararkiv i Språksamlingane etter målføregranskarar, mellom anna arkivet etter Hallfrid Christiansen og det tidlegare nemnte arkivet etter Ingeborg Hoff. I slike arkiv finst grunnlagsmateriale til publikasjonar og føredrag, brev og feltnotat som til saman gir eit interessant bilete av forskarar si gjerning og arbeidsmåtar (jf. Jansson 2018).

## Norrønt

Dei norrøne samlingane består av ein del administrativt materiale frå Gammalnorsk Ordboksverk, ei omfattande samling ordbokssetlar og diplomavskrifter som har vore brukt til dokumentasjon av det norske i norrøn og mellomnorsk tid (jf. Ebeling m.fl. 2014: 6–7 og vedlegg 1). Setelmaterialet er basert på litterære kjelder frå den klassiske norrøne (700–1350) og mellomnorske (1350–1525) perioden. Diplommaterialet er avtaletekstar og liknande frå norrøn og mellomnorsk tid. I tillegg råder Språksamlingane òg over digitaliserte utgåver av Johan Fritzner (1886–1896) *Ordbog over det gamle norske Sprog*, Richard Cleasby & Gudbrand Vigfusson (1874) *An Icelandic-English Dictionary*, samt Geir T. Zoëga (1910) *A Concise Dictionary of Old Icelandic*. Fritzners ordbok inneheld ordforklaringar og relevante sitat frå norrøne kjelder. Cleasby & Vigfusson og Zoëga er begge engelskspråklege ordbøker av relevans for for-

sking og undervisning i norrønt. Ordbøkene finst i samsøk i ei tidleg testutgåve (<https://oda.uib.no/ordbok/?men=norrone&mc0=fritzner&mc1=clevig&mc2=zoega>).

Setelsamlinga frå Gammalnorsk Ordboksverk er for ein stor del innlemma i *Menota – Medieval Nordic Text Archive* ([www.menota.org](http://www.menota.org)) i løpet dei siste tiåra. Det fellesnordiske Menota-prosjektet med 15 deltakande institusjonar er eit digitalt arkiv der stadig fleire norrøne tekstar blir lagde til. Tekstane er transkribert (diplomatarisk nivå), og mange er også tilgjengeleg på faksimilært nivå og transkribert til normalisert nivå. Nokre av tekstane er lemmatisert, og ei gruppe av desse er dessutan morfologisk analysert. Her skal også dei digitaliserte utgåvene av *Diplomatarium Norvegicum* som er på norrønt innlemmast. I Juliane Tiemanns artikkel i dette nummeret blir delar av ordbokssetel-samlinga brukt som grunnlag for å undersøke den syntaktiske variasjonen ein finn i gammalnorsk, og dei faktorane som påverkar ulik ordstilling i teksten til *Konungs skuggsjá* (Kongespeilet) etter AM 243 b α fol.

## Framtida for Språksamlingane – ein levande infrastruktur for norsk språk

Dei samlingane som blei overført frå Universitetet i Oslo, oppfyller ikkje moderne krav til (digital) synlegheit og bruk som opne, utvekslingsmoglege data. Dette er ei av dei store arbeidsoppgåvene som Språksamlingane ved UiB arbeider med i dag. UiB, og særleg Universitetsbiblioteket, har utropt opne data til eit strategisk satsingsområde. Dette betyr populært sagt at alt skal vere offentleg tilgjengeleg for kven som helst å bruke og etterprøve. For arkiv og samlingar utgjer dette ei særleg utfordring, då digital synleggjering og open utveksling av data er ressurstungt å implementere, særleg på IKT-sida.

Språksamlingane tek altså mål av seg å bli ein transparent, open forskingsressurs, samstundes som samlingane har spesifikke samfunnsoppdrag å utføre (jf. Mykings artikkel i dette nummeret). Dette betyr at samlingane både arbeider innanfor det handlingsrommet som rammeløypingar gir, og tek del i eksternt finansierte prosjekt, der ein kan få utvikla og utført nokre av dei oppgåvene som manglar i oppgraderinga av Språksamlingane. Slik det har vore til no, har størstedelen av moderniseringa på systemsida, dei underliggande databasane og visningssidene,

lege innanfor Språksamlingane sine egne løyvingar. Modernisering på datasida har lege både i prosjektregi og innanfor rammeløyvinga.

Brukbarheit er lett å tale om som omgrep, men kan vere vanskeleg å oppnå. For det første må det bli gjort klart kven infrastrukturen skal vere brukbar for. Dei fleste vil nok utan vidare seie at Språksamlingane skal vere ein ressurs for språkforskarar i universitets- og høgskulemiljøet. Men i realiteten går Språksamlingane langt ut over dette i faktisk bruk. Ordboksressursane er brukt av heile det norske folket, og særleg dei i ein læringsfase – skuleelevar og studentar. Stadnamnressursane blir brukt av forvaltarane av Lov om stadnamn, namnegranskarar, samt historikarar og lokalhistorisk interesserte. Termportalen og dei fagspråklege ressursane blir brukt av dei miljøa der fagspråk er i bruk og innanfor universitetsverda. Målføreresressursane og dei norrøne ressursane blir derimot i hovudsak brukt av forskarar og berre i mindre grad av folk utanfor akademia.

Å skape ein brukbar forskingsinfrastruktur for eit enkelt fagområde kan vere ganske vanskeleg, men Språksamlingane har den fordelan at dei har relevans – og er nødvendig – for store delar av samfunnet, samstundes med at dei er ein forskingsinfrastruktur for språkforskingmiljøet. Dette krev eit særleg fokus på utvikling av samlingane for å kunne både skape meningsfylte ressursar for forskarar, forvaltarar og lekfolk og samstundes utbreie kunnskap og bruk av forskingsinfrastrukturen til andre og nye fagområde og interesserte.

Det ligg store utfordringar for Språksamlingane ute i horisonten – men òg store moglegheiter. Vi får gripe dei.

## Omtalte nettressursar

*Arkivportalen, Språksamlinganes sorterte arkiv.* Arkivportalen.

<https://www.arkivportalen.no/search?repository=SPR> (henta 28.9.2022).

*Bokmålsordboka.* Språkrådet og Universitetet i Bergen. <http://ordbokene.no> (henta 28.9.2022).

*Corpuscle.* Universitetet i Bergen. <https://clarino.uib.no/korpuskel/corpus-list> (henta 28.9.2022).

*Digitale norrøne ordbøker i Språksamlingane.* Universitetet i Bergen.

<https://oda.uib.no/ordbok/?men=norrone&mco=fritzner&mc1=cl>

- evig&mc2=zoega (henta 28.9.2022).
- Language Infrastructure made Accessible (LIA)*. Tekstlaboratoriet, Universitetet i Oslo. <http://www.tekstlab.uio.no/LIA/> (henta 28.9.2022).
- Lokalnamn* [frå Rogaland]. Egil Sølvsberg. [www.lokalnamn.no](http://www.lokalnamn.no) (henta 28.9.2022).
- Medieval Norse Text Archive*. Stiftelsen Menota. [www.menota.org](http://www.menota.org) (henta 28.9.2022).
- Målføresynopsisen* [nytt digitalt søkegrensesnitt]. Universitetet i Bergen. <https://synopsis.uib.no>
- Målføresynopsisen og Norsk Dialektatlas, innskanna bilete*. Universitetet i Bergen. <https://www.uib.no/ub/114703/målføre-og-dialektologi> (henta 28.9.2022).
- Norsk Ordbok*. Universitetet i Bergen. [www.no2014.uib.no](http://www.no2014.uib.no) (henta 28.9.2022).
- Norske stedsnavn | Norske stadnamn*. Universitetet i Bergen. <https://toponymi.spraksamlingane.no> (henta 28.9.2022).
- Nynorskordboka*. Språkrådet og Universitetet i Bergen. <http://ordbokene.no> (henta 28.9.2022).
- Spesialglimt frå Språksamlingane*. Universitetet i Bergen. <https://pahoyden.no/tag/språksamlingane> (henta 28.9.2022).
- Språksamlingane*. Universitetet i Bergen. [www.språksamlingane.no](http://www.språksamlingane.no) (henta 28.9.2022).
- Termportalen*. Universitetet i Bergen. <https://term.uib.no> (henta 28.9.2022).

## Litteratur

- Andersen, Gisle (red.). 2012. *Exploring Newspaper Language. Using the web to create and investigate a large corpus of modern Norwegian*. Amsterdam: John Benjamins.
- Andersen, Gisle og Marita Kristiansen. 2015. Termportalen som infrastruktur for terminologi i Norge. *Terminologien*, 53–60.
- Andersen, Gisle og Peder Gammeltoft. Under publisering. The role of CLARIN in advancing terminology: The case of Termportalen – the national terminology portal for Norway. [Blir publisert i antologi om CLARIN ERIC]

- Cleasby, Richard og Gudbrand Vigfusson. 1874. *An Icelandic-English Dictionary*. Oxford: Clarendon Press.
- Ebeling, Signe Oksefjell, Kristian E. Kristoffersen, Øystein Vangsnes og Åse Wetås. 2014. *Den forskings- og undervisningsmessige verdien av eit utval samlingsressursar ved ILN. Rapport frå ei arbeidsgruppe nedsett av instituttleiari ved ILN hausten 2014*. Oslo. Tilgjengeleg frå: [https://www.uib.no/sites/w3.uib.no/files/attachments/d-sak-1-2-samlet\\_samlingsrapport\\_o.pdf](https://www.uib.no/sites/w3.uib.no/files/attachments/d-sak-1-2-samlet_samlingsrapport_o.pdf)
- Fritzner, Johan. 1886–1896. *Ordbog over det gamle norske Sprog – Omarb., forøget og forbedret* Udg. 1–3. Kristiania: Den norske Forlagsforening.
- Gammeltoft, Peder. 2022. *Myking – hvad er det for et navn? Et indblik i norske ing-navne*. *Bergen Language and Linguistics Studies* 12 (1): 39–69. <https://doi.org/10.15845/bells.v12i1.3534>.
- Gammeltoft, Peder. Under publisering. Placename types and their distribution – what do they signify? [Blir publisert i *UBAS – Universitetet i Bergens arkeologiske skrifter*.]
- Hoff, Ingeborg. 1960. Norsk dialektatlas. *Norsk tidsskrift for sprogvidenskap* XIX, 595–622.
- Jansson, Benthe Kolberg. 2018. På sporet av Ingeborg Hoff. I *Aust i Vika. Årbok for Østfold Mållag*, red. M. Aasbrenn, 20–42. Fredrikstad: Kvitveis forlag.
- Kristiansen, Marita. 2014. 11. Concept change, term dynamics and culture-boundness in economic-administrative domains. I *Dynamics and Terminology: An interdisciplinary perspective on monolingual and multilingual culture-bound communication*, red. Rita Temmerman og Marc Van Campenhoudt, 235–256. Amsterdam: John Benjamins.
- Kunnskapsdepartementet. 2015. Statsbudsjettet for 2016 kap. 260 - Tildelingsbrev for Universitetet i Bergen. Henta frå <https://www.regjeringen.no/contentassets/d5cc2f23a57040b792abo6100f57fofo/tildelingsbrev-uib-2016-1997743.pdf>
- Lapponi, Emanuele, Martin G. Søyland, Erik Velldal og Stephan Oepen. 2018. «The Talk of Norway: a richly annotated corpus of the Norwegian parliament, 1998–2016.» *Language Resources and Evaluation* 52, 873–893.
- Lyse, Gunn Inger. 2021. Ut med *adamsslekt* og inn med *arveprinsesse*? Leksikografiske metodar i revisjonen av *Bokmålsordboka* og *Nynorskordboka*. *Nordiske Studier i Leksikografi* 15, 215–224. Henta frå <https://tidsskrift.dk/nsil/article/view/124023>

- Meurer, Paul. 2012. Corpuscle – a new corpus management platform for annotated corpora. I *Exploring Newspaper Language. Using the web to create and investigate a large corpus of modern Norwegian*, red. Gisle Andersen, 29–50. Amsterdam: John Benjamins.
- Vasshus, Krister. Under publisering. Korleis og i kor stor grad har den vikingtidige kontakten på kryss og tvers av Skagerrak sett spor etter seg i stadnamnutvalet? [Blir publisert i antologi/konferanserapport etter *Den sjuttonde nordiska namnforskarkongressen*, Helsingfors 2021.]
- Wetås, Åse. 2014. «Nordisk dialektkorpus og Norsk målføresynopsis – ei samanlikning av to kjelder til talemålsforskning frå kvar si tid.» I *Språk i Norge og nabolanda. Ny forskning om talespråk*, red. Janne B. Johannessen og Kristin Hagen, 315–335. Oslo: Novus.
- Zoëga, Geir T. 1910. *A Concise Dictionary of Old Icelandic*. Oxford: Clarendon Press.

Peder Gammeltoft  
Universitetet i Bergen  
Universitetsbiblioteket  
Postboks 7808  
5020 Bergen  
peder.gammeltoft@uib.no

Randi Neteland  
Universitetet i Bergen  
Institutt for lingvistiske, litterære og  
estetiske studier  
Postboks 7805  
5020 Bergen  
randi.neteland@uib.no



# Frå “språksamlingane” til *Språksamlingane* – prosessen og vurderingane

Av Johan Myking

I 2014 sa Universitetet i Oslo frå seg ansvaret for dei nasjonale språksamlingane, og ein omfattande offentleg debatt følgde. I 2015 overførte Kunnskapsdepartementet ansvaret for samlingane til Universitetet i Bergen med ei rammeløyving på seks millionar kroner. Artikkelen skildrar korleis Universitetet i Bergen posisjonerte seg og dei ulike strategiske vurderingane bak prosessen. Forsøket på redningsaksjon gjennom ei nasjonal ordbokgruppe med representasjon frå dei norske universiteta vert òg presentert. Universitetet i Bergen har akseptert ansvaret for samlingane som eit nasjonalt oppdrag, og slår no saman samlingane med eigne samlingar og samlingar frå andre institusjonar. Eit viktig resultat er at institusjonen har fått eit aktivt leksikografisk fagmiljø med fagleg ansvar for Bokmålsordboka, Nynorskordboka og Norsk Ordbok og tilleggsfinansiering over statsbudsjettet. Den avgrensa rammeløyvinga set grenser for drifta av andre delar av samlingane. Målet er framleis å utvikla samlingane som nasjonal forskingsressurs.

## 1 Innleiing

Universitetet i Bergen fekk gjennom hausten 2016 mange medieoppslag på flyttinga av dei store språksamlingane frå Universitetet i Oslo. Materialet omfatta både digital og fysisk språkdokumentasjon, og særleg den siste typen viste seg å vere medievennleg: Det som i utgangspunktet var 60 tonn og 15 trailerlass med ord og namn, voks vel snøgt til 70 tonn fordelte på 16 lass. (Kyrkjebø og Myking 2017)

Vedtaket om avvikling av språksamlingane ved Universitetet i Oslo i 2014 skapte eit omfattande offentleg ordskifte og stor kulturpolitisk uro for at denne verdfulle språklege dokumentasjonsarven skulle forvitra.

Universitetet i Bergen fekk i november 2015 i oppgåve av Kunnskapsdepartementet å overta ansvaret for samlingane. Det er enno for tidleg å skriva den fulle historia om denne flyttinga og bakgrunnen for den, sjølv om hovuddraga er velkjende. Kjeldematerialet er stort, men ikkje fullt utnytta her. Samstundes var denne artikkelforfattaren aktør i prosessen som leiar for det instituttet (LLE) som tok mot samlingane. Når han no er si eiga kjelde til delar av prosessen, er framstillinga derfor sjølvstendig og sett berre frå ei av sidene, UiB, sin synsstad. Hovudvekt ligg på den politiske delen av prosessen sommaren 2014 og desember 2015, då det overordna problemet gjekk ut på å finna ei nasjonal løysing. Etter flyttinga er det no gått fem år med utviklings- og utvidingsarbeid, og samlingane er i dag etablert som ein nasjonal ressurs for utforskinga av norsk språk ved Universitetet i Bergen under namnet Språksamlingane (sjå Gammeltoft og Netelands artikkel).

## 2 Historisk bakgrunn

Oppbygginga av vitenskaplege dokumentasjonssamlingar heng nøye saman med utviklinga av ein norsk språkvitenskapleg tradisjon frå 1800-talet og frametter, men var særleg omfattande gjennom den første delen av 1900-talet. Norsk Målførearkiv vart grunnlagt ved Universitetet i Oslo i 1936. Norsk Stadnamnarkiv vart skipa av Kyrkjedepartementet i 1921, og overført til Universitetet i Oslo i 1978. Mellom 1939 og 1944 låg det i Bergen – overført til Bergen museum under Gustav Indrebø, men returnert i 1944 av Per Hovda under sterk protest frå leiinga ved museet (Helleland 1997, Ådland 1997). Norsk Leksikografisk Institutt vart i 1972 lagt til Universitetet i Oslo med nasjonalt ansvar for dei tre ordbokverka (bokmål, nynorsk og gammalnorsk). Det samla materialet, både språklege datasamlingar og anna språkvitenskapleg forskingsmateriale, har vorte omtala som “språksamlingane ved Universitetet i Oslo”. I dag utgjer desse overførte samlingane stammen i ei organisatorisk eining med namnet Språksamlingane ved Universitetet i Bergen.

I Bergen skipa Torleiv Hannaas Bergens Museums Folkeminnesamling i 1921. Frå den organiserte Gustav Indrebø frå 1931 dei såkalla skulebarnoppskriftene, som seinare la grunnlaget for stadnamnsamlinga ved Nordisk institutt. Olai Skulerud grunnla litt seinare Målføresamlinga, og begge samlingane har vore ein del av Det humanistiske fakultet ved



Universitetet i Bergen fram til i dag gjennom faget nordisk. I dag er alle desse samlingane ved UiB samanslegne med dei nye frå Oslo og andre samlingar, og ligg altså under Språksamlingane.

Prosjektet Norsk Ordbok vart sett i gang i 1930 av Det Norske Samlaget, og Universitetet i Oslo vart den vitskapelege samarbeidspartnaren, med nemnda for Norsk Ordbok som samarbeidsorgan frå 1972. Ordbokverket skulle etter planen ferdigstillast i 2014, men Universitetet i Oslo endra haldning til vidareføringa av arbeidet, og avslutningsfasen vart turbulent (Vikør og Wetås 2016: 32–34). Aktiviteten ved samlingseiningane var i praksis avvikla tidlegare.

Institutt for lingvistikk og nordistikk (ILN) ved Universitetet i Oslo sette hausten 2013 ned ei arbeidsgruppe for å vurdere samlingane som forskings- og undervisningsressurs. Drifta av samlingane – rekna til 19 millionar kroner årleg – vart frå ca 2010 rekna som ikkje finansiert, men som ein del av instituttøkonomien (Vikør og Wetås 2016: 32). På grunnlag av rapporten frå arbeidsgruppa vedtok styret for ILN derfor 10.6.2014 mot to røyster å avvikla samlingane og leita etter andre interessantar som kunne overta, eventuelt få materialet deponert i samsvar med gjeldande statlege retningsliner (ILN møte nr 3/2014 V-SAK 4, med rapporten frå arbeidsgruppa vedlagd). Trass i det omfattande ordskiftet som kom det neste halvåret, endra ikkje instituttet eller Universitetet i Oslo som institusjon denne posisjonen (sjå t.d. UiO/KD Status).

Desse ytre sakstilhøva er offentleg kjende frå før (sjå òg t.d. saksframstillinga i Norsk ordbokpolitikk, 2014). Nedleggingssaka skapte derimot ein del problem for dei mange instansane som hadde interesser og behov knytt til samlingane, derfor vart dette definert som eit nasjonalt problem som måtte løysast. Bakgrunnen for nedleggingssaka ved Universitetet i Oslo og dei prosessane som førte frå til den, er komplekse, og det er til no ikkje gjeve noka samla framstilling av dei. Heller ikkje denne artikkelen kan gje ei slik samla framstilling, men må avgrensast til å skildra eit avgrensa utsnitt av den samla saga.

### 3 Dei ulike interessentane og deira særlege behov

Språkrådet sit med forvaltingsansvaret for dei norske standardordbøkene – Bokmålsordboka og Nynorskordboka. Fram til 2016 delte Språkrådet eigarskapet til desse ordbøkene med Universitetet i Oslo gjennom ILN.

Språkrådet lagar ikkje ordbøker sjølv, og ville derfor få bruk for ein ny fagleg samarbeidspartnar dersom Universitetet i Oslo skulle seia frå seg dette ansvaret. Det Norske Samlaget og dei tilsette ved Norsk Ordbok var på si side ottefulle for at ferdigstillinga av den digitale versjonen og revisjonen av alfabetstrekket a–h skulle stansa. Arkivet til Norsk Riksmålsordbok var del av språksamlingane, og omsynet til leksikografien for bokmål vart òg på ymse vis trekt inn i ordskiftet. Ikkje minst var Språkrådet uroleg for lagnaden til dei store stadnamnsamlingane, som er eit viktig dokumentasjonsmateriale for namnekonsulenttenesta. Uroa for namnesamlingane hang (og heng) nøye saman med den allmenne uroa for disiplinen namnegransking, som har opplevd kraftig nedgang ved dei norske universiteta, m.a. stod fleire vitskapleg tilsette ved Universitetet i Oslo framfor pensjonsavgang då denne saka vart aktuell. I språksamlingane var det òg lagra store mengder anna forskingsmateriale og språkleg datamateriale, og materialet i Norsk målførarkiv hadde tilknytning både til Norsk Ordbok og til den språkvitskaplege dialektforskninga meir allment.

Nasjonalbiblioteket vart av ILN peika ut som den naturlege mottakaren av materialet. Frå andre hald, særleg tydeleg uttrykt av Språkrådet i eit notat om ordbokpolitikk (nedanfor, s. 8) vart det innvendt at Nasjonalbiblioteket neppe ville gå inn som aktivt ordbokmiljø, derfor ville ein aktiv forskingsinstitusjon utanfor eigarskapet til Kulturdepartementet vera ei nødvendig løysing. I fagmiljøet ved Universitetet i Bergen, ved Institutt for lingvistiske, litterære og estetiske studium (LLE), rådde det tilsvarande otte for at lagring ved Nasjonalbiblioteket ville føra materialet på for lang avstand til den aktive språkforskninga.

I opptakten til prosessen gav Språkrådet våren 2014 ei utgreiing til Kulturdepartementet om desse problemstillingane, og gjorde m.a. framlegg om at

det opprettes et senter for vitenskapelig dokumentasjon av de språka Norge har ansvar for, som har som kjerneoppgave å vedlikeholde og utvikle norske språksamlinger fra norrønt til moderne norsk, og dermed føre videre det arbeidet som Norsk leksikografisk institutt ble tillagt i 1972, og som har pågått ved Universitetet i Oslo til nå (Norsk ordbokpolitikk, 2014, s. 6)

Grunntanken om eit uavhengig og nasjonalt dokumentasjonssenter vart òg eit viktig element i diskusjonane, m.a. i den nasjonale ordbokgruppa

(sjå nedanfor). Interessene sprikte altså litt, men samla seg etter kvart i ønsket om at språksamlingane burde liggja ved eit universitet og vera ein levande språkvitskapleg ressurs drifta og utvikla av eit aktivt forskingsmiljø.

#### 4 Tidslinja var om lag slik

Tanken om at Universitetet i Bergen kunne overta samlingane, vart lansert offentleg av Helge Sandøy i universitetet si internavis På Høyden (4.8.14) Ei veke tidlegare hadde Gunnstein Akselberg lansert ein tanke om at Universitetet i Bergen åleine burde satsa på namnegransking og -forvaltning, men utan å dra det så langt som til flytting av materialet frå Oslo (Akselberg 2014). Ideen om flytting fekk støtte på leiarplass i Universitetet i Oslo si avis Uniforum (22.8.14) og vart lufta på møte i universitetsstyret 28.8.14. Rektor Dag Rune Olsen er referert slik:

c) [Oddny] Miljeteig ba også om status i saken med en eventuell overtagelse av Språksamlingene fra Universitetet i Oslo (UiO). Rektor orienterte om at UiB er interessert i å overta samlingene under noen gitte forutsetninger og at det pågår en dialog med UiO om dette (UiB 71/2014)

Den same kvelden, 28.8.14, var det innkalla til eit stort offentleg møte om denne saka på Litteraturhuset i Oslo. Haldninga til rektor ved Universitetet i Bergen vart referert der, og dermed kan ein seia at Universitetet i Bergen som interessent var tydeleg presentert. Spørsmålet om ein annan type finansiering enn vanlege resultatinntekter vart òg på dette tidspunktet tydeleg markert (Uniforum 1.9.14)

Etter initiativ og under samordning frå Nynorsk kultursentrum ved direktør Ottar Grepstad vart det skipa ei nasjonal ordbokgruppe med representantar frå sju universitet og høgskular. Grappa hadde 14 referatførte telefonmøte mellom oktober 2014 og desember 2015. Instituttleiaren ved LLE (altså denne artikkelforfattaren) representerte Universitetet i Bergen i denne grappa og kunne derfor fungera som bindeledd mellom grappa og universitetsleiinga. Rektor Per Halse ved Høgskulen i Volda deltok òg i grappa. Nynorsk kultursentrum la ned eit stort arbeid i og rundt ordbokgrappa, særleg gjennom fleire viktige dokumentasjonsnotat.

Det vart etter kvart klårt for alle at ILN ikkje kom til å endra standpunkt. I staden tok ILN kontakt med andre forskingsinstitusjonar for å sondera interessa for samlingane. Sonderingsmøtet med Universitetet i Bergen (dvs. LLE) vart halde i desember 2014. Etter møtet laga LLE på oppmoding eit notat som greidde ut vilkår, risiko og mogelege vinstar ved å overta samlingane (brev til ILN 10.2.15).

Den 27. april 2015 hadde rektor Dag Rune Olsen fått i stand eit møte med den politiske leiinga i Kunnskapsdepartementet, representert ved statssekretær Bjørn Haugstad. Rektor Per Halse ved Høgskulen i Volda deltok på møtet. Det gjorde òg instituttleiaren ved LLE og dåverande dekan for Det humanistiske fakultet, noverande rektor Margareth Hagen frå Universitetet i Bergen.

Frå Kunnskapsdepartementet si side vart det peika på at institusjonane har rett til å gjera egne faglege prioriteringar som departementet ikkje overprøver. Det kom likevel tydeleg fram at departementet var oppteke av å finna ei ordning med Kulturdepartementet slik at òg interessene til dette departementet kunne varetakast. Kunnskapsdepartementet trong òg å undersøkjia med Universitetet i Oslo for å finna ut kva som eigentleg skulle avleverast av materiale og kva kostnader dette representerte. Møtedeltakarane frå Bergen og Volda vart lova at ingen avgjerder skulle takast utan at dei to institusjonane vart konsulterte (På Høyden 29.4.15). Miljøet i Bergen hadde likevel fått eit klårt inntrykk av at Kulturdepartementet delte standpunktet frå Universitetet i Bergen om at dei økonomiske ressursane måtte følgja med samlingane ved flytting til Bergen, i strid med haldninga til Universitetet i Oslo (På Høyden 10.4.15). Ved utgangen av april rådde det såleis ei varsamt optimistisk stemning.

Mellom dette møtet i slutten av april og fram til midten av november 2015 gjekk sjølv sagt ordskiftet vidare, og særleg ordbokgruppa og Nynorsk kultursentrum gjorde ein intens innsats med å mobilisera støtte i det politiske miljøet, først og fremst retta mot Stortinget. Likevel var det nok helst den tosidige kontakten mellom universitetsleiinga ved Universitetet i Bergen og den politiske leiinga i Kunnskapsdepartementet som frå då av dreiv prosessen framover, og den kontakten er ikkje dokumentert i detalj. Det kom ikkje ut særleg mykje informasjon om framdrifta i saka, og så seint som november 2015 verka saka uavklåra og kaotisk. Ho vart teken opp i Stortinget, det var mellom anna strid om brukarrettleiinga for Norsk Ordbok, som Kulturdepartementet hadde løyvd midlar

til, men som Universitetet i Oslo ikkje såg det mogeleg å fullføra (På Høyden 11.11.15).

Tidleg i november 2015 nådde det likevel fram eit rykte til miljøet i Bergen om at saka hadde løyst seg, og 23. november vart dekan og instituttleiar innkalla til rektor og orienterte om at Universitetet i Bergen dagen før hadde fått i oppdrag å overta samlingane. I tildelingsbrevet er oppdraget formulert slik:

Universitetet i Oslo ønsker ikke å ha ansvaret for videre drift og forvaltning av samlingene knyttet til leksikografi og navnegranskning. Etter en samlet vurdering har departementet kommet til at samlingene overføres til Universitetet i Bergen. Budsjettrammen til Universitetet i Bergen er dermed økt med 6 mill. kroner. Universitetene må vurdere om overføringen vil kunne ha arbeidsrettslige konsekvenser for ansatte. Universitetet i Bergen bør vurdere eventuelle samarbeidspartnere i den videre drift, forvaltning og bruk av samlingene. (Tildelingsbrev 2016, s. 8)

Seks millionar var monaleg mindre enn ein hadde vona på, men frå då av tok sjølve flytteprosessen til. Den fekk sin eigen dynamikk, og varte i første omgang til sommaren 2017. Då vart materialet omorganisert reint fysisk, det vart oppnemnt ei styringsgruppe direkte under universitetsleiinga, og det vart rekruttert fagleg leiar, ordbokredaktør og arkivar. Det vart i april 2016 halde eit nasjonalt arbeidsseminar i Bergen som start på utviklinga av samlingane, og fagmiljøa ved Universitetet i Bergen, Høgskulen i Volda, NLA Høgskolen og Høgskulen på Vestlandet ordna dei neste par åra tre ulike seminar i Bergen og Volda for å greia ut korleis samlingane kunne utnyttast i lærarutdanningane. Saka utløyste altså mykje energi og dugnadsvilje (På Høyden 19.4.16).

Både ved Universitetet i Bergen og Universitetet i Oslo var fagmiljøa nøye med å skilja mellom språksamlingane og dei ordbokprosjekta som bygde på samlingane. Likeins var dokumenta som Nynorsk kultursentrum leverte, stringente og tydelege i å få fram skiljet. For andre aktørar kan nok skiljet i utgangspunktet ha vore temmeleg uklårt og abstrakt, men alvoret seig tidleg inn over dei fagleg ansvarlege. I ordlegginga “samlingene i leksikografi” i oppdragsbrevet låg det eit tydeleg krav om at dei ufullførte prosjekta skulle følgjast opp, særleg digitaliseringa av den første delen av Norsk Ordbok (a–h). Så tidleg som i 2015 tok direktøren ved Nynorsk kultursentrum grep saman med rektor ved Høgskulen i Volda for å få fullført brukarrettleiinga for ordbokverket, som Kultur-

departementet alt hadde løyvd midlar til, og brukarretteiinga vart fullført i 2016 (Vikør 2018).

Tildelingsbrevet nemner eksplisitt det nasjonale ansvaret for leksikografi og namnegransking. Universitetet i Bergen og Språkrådet skreiv i 2016 under på ein avtale om delt eigarskap til standardordbøkene. Med grunnlag i denne avtalen fekk Språkrådet så finansiering over kulturbudsjettet til å få revidert standardordbøkene over fem år med prosjektstart i 2018. Frå 2019 kom det òg startløyving til revisjonen og digitaliseringa av Norsk Ordbok. Dette revisjonsarbeidet går no føre seg med finansiering over statsbudsjettet med ti års tidshorisont, 2019–2029 (prosjektet NO-AH). I 2018 formaliserte Det Norske Samlaget, Høgskulen i Volda og Universitetet i Bergen eit felles ansvar for Norsk Ordbok. Høgskulen i Volda står som medsøkjjar om løyvingar til prosjektet NO-AH, og den særskilde nemnda for Norsk Ordbok, som hadde eksistert sidan 1972, vart reetablert gjennom denne avtalen. På dette punktet har altså Universitetet i Bergen følgd opp formuleringa i tildelingsbrevet om å “vurdere eventuelle samarbeidspartnere” i drifta av samlingane.

Samla sett er desse tilføyningane med på å sementera *Språksamlingane* som ein nasjonal ressurs for utforskinga av norsk språk. Den samla aktiviteten i 2021 tel rundt rekna 12–13 årsverk og har eit omfang på om lag 19 millionar. Frå 2021 er opprettinga av Termportalen medrekna som ny aktivitet (jf. Prop 108 L og Språkstatus 2021).

## 5 Kva stod på spel for fagmiljøet ved UiB?

ILN møtte som nemnt LLE til diskusjon om ei eventuell overføring i desember 2014, og bad då om ei utgreiing frå LLE om saka. Posisjonen til LLE vart formulert slik:

LLE gjer ikkje krav på å overta heile porteføljen av samlingar, heller ikkje har instituttet noko krav om at alt som finst av språksamlingar i Noreg skal lokaliserast i Bergen. Det er dei noverande samlingane ved Universitetet i Oslo dette notatet handlar om. Å ta vertskapsansvaret for ein ny organisasjon er ei markering både av at institusjonen tek nasjonalt ansvar og ønskjer eit samarbeid med alle andre interesserte forskingsmiljø og institusjonar i landet. Dette inkluderer òg Nasjonalbiblioteket og Språkbanken, som har vore framme i debatten. Ved å sjå den nye organisasjonen som del av ei ”klynge”

kunne ein sikra at samlingane dannar eit nasjonalt tyngdepunkt for dette feltet (LLE/ILN 10.2.15)

I dette låg det ei offensiv innstilling balansert mot eit ønske om å framstå som inkluderande, men òg ulike former for gardering. Saka var på det tidspunktet ikkje drøfta formelt med noko styrande organ ved Universitetet i Bergen, og vart heller aldri nokon gong behandla som vedtakssak. Det var viktig for instituttleiinga ved LLE (dvs. denne artikkelforfatteren) å få fram dei risikomomenta som fanst og samstundes tona ned det inntrykket som kanskje var skapt av at fagmiljøet i Bergen stilte seg unisont entusiastisk til overtakinga.

Dei etablerte samlingane ved HF/Universitetet i Bergen hadde i lang tid vore inne i ei negativ utvikling. Drifta hadde lenge vore ordna med samlingsstyrarar frå den vitskaplege staben som kunne bruka 25 % av arbeidstida si på samlingane, i tillegg til fast tilsette arkivarar. Denne ordninga hadde i mange år vore under press på grunn av ressursituasjonen ved fakultetet. Styraren for Stadnamnsamlinga hadde gått av med pensjon i 2012, og stillinga var inndregen. Så tidleg som hausten 2014 fekk instituttet klår melding om at det ikkje ville vera aktuelt å reetablere denne stillinga, sjølv ikkje som premiss for å ta mot samlingane frå Oslo. Også samlingane ved andre institutt opplevde liknande problem, og i fakultetet hadde det lenge vore diskusjonar om å overføra samlingane til Universitetsbiblioteket (UBB). Fagmiljøet såg det slik at dette ville vera å avslutta samlingane som “levande” forskingsressursar (t.d. Akselberg 2014, sjå òg På Høyden 23.5.19).

Etter faglege initiativ frå fleire miljø (m.a. lesarinnlegg i På Høyden, Bang 2014) sette derfor HF-fakultetet i ned ei arbeidsgruppe i april 2015 for å greia ut framtida for samlingane ved HF/Universitetet i Bergen. Diskusjonen om flytting av språksamlingane frå Oslo var då i gang, men den saka var uavklåra og førebels ein nokså usikker premiss for arbeidsgruppa. Og omvendt var altså samlingsarbeidet ved Universitetet i Bergen ein nokså usikker premiss for å kunna overta samlingane frå Oslo.

Innstillinga frå arbeidsgruppa ved HF vart levert i august og styrebehandle 20.10.15 (UiBHF Fsak 46/15). Språksamlingane frå Oslo er ikkje nemnde, og dei som var interesserte i samlingane, vart skuffa: Fakultetsstyret fann ikkje å kunna prioritera samlingane, men la til grunn at samlingane over ein periode på tre år skulle overtakast av Universitets-

biblioteket. Alle administrative ressursar som var knytte til samlingane, ville verta inndregne ved naturleg avgang.

Ein kunne sjølvstakt ikkje venta at støtta i det samansette fagmiljøet ved LLE skulle vera unison, og det gjorde ikkje stoda enklare at fagmiljøet i talemål og sosiolingvistikk i åra mellom 2017 og 2019 stod framfor nedskjering som følgje av ikkje mindre enn tre pensjonsavgangar. Det stod nok ikkje klårt for alle kollegaene kva for praktiske konsekvensar ei overtaking kunne få. Entusiasmen for å overta samlingane frå Oslo gjekk derfor hand i hand med ein viss otte og uro.

I utgangspunktet hadde tanken om Bergen fått støtte både internt og eksternt fordi ein såg det slik at samlingane då ville hamna i eit aktivt forskingsmiljø. Samlingane ville kunna syta for ei vitamininnsprøyting i språkforskinga der, og eit kompetent fagmiljø ville vera ein garantist for at samlingane kunne utviklast som levande forskingsressurs til glede for det nasjonale forskingsmiljøet. Både hausten 2014 og seinare ut gjennom 2015 kom det altså negative signal frå fakultetet til instituttet som svekte denne premissen ganske sterkt.

Instituttleiinga var på mange måtar meir aktivt interessert i saka enn delar av fagmiljøet, og såg det slik at vurderinga av risikoen måtte kombinerast med ei tilsvarende vurdering av vinstpotensialet. I ei tid med inndragingar og nedgang i dei disiplinane som låg til grunn for samlingane, kunne språksamlingane frå Oslo oppfattast som den einaste voksterimpulsen det var mogeleg å sjå føre seg i både nær og noko fjernare framtid, derfor burde ein ikkje la denne sjansen gå frå seg. Det låg framleis føre ein premis om at det skulle følgja økonomiske ressursar med samlingane, og gjennom heile prosessen fram til november 2015 var det eit tal på oppunder 20 millionar kroner ein vona på.

## 6 Felles eigarskap eller nasjonalt oppdrag?

Etter initiativ frå Nynorsk kultursentrum vart det altså tidleg sett ned ei nasjonal ordbokgruppe, som mellom 31.10.14 og 19.12.15 hadde 14 referatførte møte. Gruppa la ned mykje arbeid og engasjement mot det politiske miljøet, særleg mot stortingsmiljøet og Kulturdepartementet. Eit omfattande dokumentasjonsmateriale vart utvikla og samordna av direktøren, først og fremst brev til Kulturdepartementet 27.11.14 om ny driftsorganisasjon (NK 1) med vedlagd kostnadskalkyle (NK 2). Dette



vedlegget var ein omfattande analyse av korleis ein ny driftsorganisasjon for samlingane og ordbøkene kunne sjå ut, med ein tilsvarende grundig kostnadskalkyle.

Med litt svingande tal enda kalkylen frå Nynorsk kultursentrum på rundt 5 millionar for ordbøkene og 12 millionar for samlingane i årlege driftsutgifter, målt i 2016-kroner. Til samanlikning hadde ILN rekna desse utgiftene til om lag 19 millionar kroner årleg (ILN møtenr 3/2014 V-SAK 4). Notatet rådde til felles samlokalisering ved Universitetet i Bergen, eventuelt ei delt løysing mellom Universitetet i Bergen (samlingar) og Høgskulen i Volda (ordbøker). I notatet låg det òg eit synspunkt som av mange vart oppfatta som kontroversielt – at samlingane og ordbøkene burde lokaliserast i eit solid nynorsk miljø.

Det vart òg sendt brev til Kulturdepartementet (NK 3), og det same innhaldet kom i kronikkform i Bergens Tidende (Solholm 2015). Både ordbokgruppa og Nynorsk kultursentrum gjorde altså sitt til at tanken om Bergen som lokaliseringsstad var godt og grundig marknadsført ved årsskiftet 2014–2015.

På den andre sida var ordbokgruppa skipa og samansett ut frå eit sterkt ønske om sameigarskap og delt ansvar mellom dei norske forskingsinstitusjonane. Alle dei dåverande universiteta var representerte i gruppa. Gruppa samla seg om at samlingane kunne verta eit nasjonalt fellesføretak etter § 1-4.4 i lov om universitet og høgskular, om lag etter modell av Artsdatabanken. Den ligg lokalisert ved NTNU utan å vera ein organisatorisk del av universitetet. Denne tanken vart LLE sin posisjon, m.a. overfor ILN og i drøftingar med universitetet sentralt, men vekte inga særleg interesse der (På Høyden 28.11.14 og brev til ILN 10.2.15).

I ettertid er det nok rett å seia at ordbokgruppa feilvurderte dei politiske straumdraga på dette punktet. Den politiske tankegangen i styringsverket går ikkje i retning nasjonale fellesføretak eller nye, uavhengige senter, men heller mot nasjonale oppdrag og nasjonalt ansvar for spesifikke felt tildelt spesifikke institusjonar. Dette vart òg løysinga i denne saka, sjølv om departementet i tildelingsbrevet òg oppmoda til å vurdere samarbeid med andre institusjonar. Norsk målførearkiv og Norsk leksikografisk institutt vart overførte til Universitetet i Oslo på 1970-talet med nasjonalt ansvar (sjå ovanfor, avsnitt 3), så i realiteten har ikkje styringsverket endra linje dei siste femti åra.

## 7 Vurderingar av prosessen

Mellom august 2014 og møtet med Kunnskapsdepartementet i slutten av april 2015 gjekk den offentlege diskusjonen føre seg med høg temperatur, samstundes som ein ved Universitetet i Bergen og i ordbokgruppa greidde ut premissane og prøvde å analysere konsekvensane, kvar for seg og i samarbeid.

Ikkje på noko tidspunkt var denne overtakinga vedtakssak i eit styrande organ ved Universitetet i Bergen, men situasjonen hadde vorte politisk. Rektor tok tidleg tak i saka etter oppspel og innspel frå dekan og instituttleiar, som begge meinte at dette reint fagleg var ei god sak for Bergen. I ettertid er det freistande å meina at det knapt hadde vore mogeleg for LLE og Universitetet i Bergen å ta nokon retrett etter møtet mellom Universitetet i Bergen og Kunnskapsdepartementet 27. april 2015. I løpet av vinteren og våren 2014–2015 hadde rektoratet ved Universitetet i Bergen teke eigarskapet til saka, og den avgjerande kommunikasjonen gjekk tovegs mellom Universitetet i Bergen og departementet. Når kandidaturet til Universitetet i Bergen og Høgskulen i Volda så ettertrykkeleg hadde blitt presentert av institusjonane sjølve, er det vanskeleg å sjå for seg at Universitetet i Bergen seinare kunne ha takka nei til samlingane. Toget var på skjenene og farten for høg. Men dette tyder òg at dei økonomiske kalkylane frå Nynorsk kultursentrum truleg ikkje spela ei avgjerande rolle i desse samtalane – ein må tru at eit tal på 17–18 millionar kroner var realpolitisk umogeleg å stå hardt på. Rektor hadde slege fast overfor departementet at aktiviteten ville verta definert og avgrensa av storleiken på rammeløyvinga, og dette vart utgangspunktet for det vidare arbeidet.

## 8 Strategiske problemstillingar på fleire nivå

Det kan vera grunn til å minna om ein del meir grunnleggjande faktorar som påverka denne prosessen.

For det første reint dei vitskaplege problemstillingane: Er arkiv og samlingar nyttige som forskingsressursar i moderne språkforskning, eller finst det meir ‘moderne’ tilnærmingar? Er leksikografisk arbeid vitskap, er det naturleg for eit forskingsuniversitet å driva med ordbokarbeid? Og for det tredje: Står det å arbeida med samlingsbasert morsmålsleksiko-

grafi i motstrid til det å sikta mot å verta eit internasjonalt toppuniversitet? Det feste seg eit inntrykk av at Universitetet i Oslo svara negativt på desse spørsmåla (jf. ILN møte nr 3/2014 V-SAK 4). Frå ordbokmiljøet tok særleg Lars S. Vikør eit kraftig oppgjer med desse haldningane (Vikør 2016). Problemstillinga vert sjølvstøtt sett på spissen når det er knytt økonomiske føringar til valet.

Det fanst òg ein del overordna universitetspolitiske føringar. For det første autonomiproblemet – institusjonane prioriterer sjølvstendig. Kunnskapsdepartementet har ikkje instruksjonsmynde og ønskjer heller ikkje å ha det. Sektoransvaret kom i tillegg: Språksamlingane var eigde av ein institusjon under Kunnskapsdepartementet, men ansvaret for bruken av dei – ordbøker og stadnamn – ligg under Kulturdepartementet, som altså ikkje kunne gripa direkte inn mot Universitetet i Oslo sjølv om det var dette departementet sine interesser som stod på spel. Det har ikkje alltid vore slik at ansvaret for språkpolitikken og ansvaret for universiteta har vore delt mellom to ulike departement, men slik har ordninga vore sidan 1990, og i denne saka kompliserte det biletet.

I åra frå 2014 vart det universitetspolitiske landskapet endra gjennom den såkalla strukturreforma i høgare utdanning. Språksamlingane og særleg Norsk Ordbok vart eit tema i meir overordna politiske samtalar mellom Universitet i Bergen og Høgskulen i Volda, og det regionale samarbeidsnettverket UH-nett Vest var òg kopla inn. Og endeleg aktualiserte redningsaksjonen frå den nasjonale ordbokgruppa spørsmålet om nasjonalt samarbeid og eigarskap for samlingane – gruppa tok til orde for ein nasjonal dugnad der alle universiteta skulle ta medansvar. Sjølv om resultatet vart annleis, var det eit viktig moment i debatten, og det samla presset styrkte nok arbeidet med å finna ei løysing.

## 9 Resultatet: Fekk ein det ein ville ha?

Leiinga ved Universitetet i Bergen avgjorde tidleg etter overtakinga at språksamlingane skulle plasserast ved Universitetsbiblioteket og skjermast for svingingane i resultatøkonomien ved fakultetet for å unngå dei problema som ILN opplevde. Dette hadde òg vore LLE sin posisjon, m.a. hadde instituttet tidleg åttvara mot å ta samlingane inn i fakultetsøkonomien (notat til rektor 26.5.15). Språkrådet hadde tidleg peika på dette problemet (Norsk ordbokpolitikk, 2014, s. 6). Organisering gjen-

nom Universitetsmuseet vart òg drege fram, men vekte ikkje interesse hjå universitetsleiinga. Dei to leksikografiske revisjonsprosjekta for ordbøkene som følgde i 2018 og 2019, vart derimot lagde til forskingsmiljøet ved fakultetet og LLE fordi dei var definerte som språkvitskaplege prosjekt. Termportalen, som vart etablert i 2021, er òg delt på den måten at den tilknytte forskarstillinga er lagd til LLE. Denne organisasjonsmodellen er ikkje særleg solid, men viser kor organisatorisk og praktisk krevjande denne saka har vore for Universitetet i Bergen.

Med ein viss rett kan det seiast at ingen av aktørane frå 2014 var i stand til å sjå heilt klårt på førehand kva dei ville få eller kva dei eigentleg bad om, og det kunne vel heller ingen krevja. Somme fekk noko anna enn dei hadde tenkt seg, og dei færreste ved Universitetet i Bergen kunne vel i 2014 sjå kva dei faktisk har fått: Eit leksikografisk fagmiljø har vakse fram etter 2018 som det største i landet. Omtalen av dette arbeidet i språkløvsproposisjonen (Prop 108 L 2020) er særst viktig som handlingsgrunnlag for åra frametter. Staten, ved Kulturdepartementet, har signalisert at det nasjonale oppdraget med ordbokarbeid inngår i den overordna strategien med å sikra statusen for norsk språk gjennom utvikling av digital språkleg infrastruktur.

Økonomisk har det på mange måtar utvikla seg heilt motsett av det aktørane såg for seg i 2015. Løyvingane over statsbudsjettet til ordbokarbeidet ved Universitetet i Bergen utgjer no om lag 9 millionar i 2022. Dette er variable prosjektinntekter, og så er altså den faste rammeløyvinga til sjølve språksamlingane 6 millionar. I kalkylen frå Nynorsk kultursentrum i 2015 var tilhøvet omvendt: rundt 12 millionar til samlingane, 5 til ordbøkene. Den smålatne rammeløyvinga set grenser for utviklingsarbeidet, særleg arbeidet med stadnamn og talemål.

Spørsmålet om kva HF-fakultetet og UBB skulle gjera med sine eigne samlingar etter overføringsvedtaket i 2015, vart ei stund liggjande uløyst, og dette valda ein viss frustrasjon (På Høyden 23.5.19). I 2020 gjorde så HF og UBB avtale om at samlingane i Bergen skulle slåast saman med språksamlingane frå Oslo og administrerast av UBB. Denne prosessen er no i gang, og både fakultetet og universitetsleiinga har løyvd av eigne midlar til dette arbeidet. Fagmiljøet i nordisk har gjeve ei klår intensjons-erklæring om interesser i samlingane (brev LLE 29.04.20). Skal samlingane vera ein ressurs særleg i utforskinga av talemål og språkending, krev det framleis at eit aktivt fagmiljø engasjerer seg aktivt i arbeidet. Her kan ein seia at fagmiljøet ved Universitetet i Bergen har fått ei ny

utfordring, fordi samlingane har hamna ved Universitetsbiblioteket og ikkje på instituttet og i fagmiljøet. Samlingane er frikopla frå instituttøkonomien, men den faglege avstanden har òg auka.

I kontakten med dei politiske styresmaktene har Universitetet i Bergen ved fleire høve mellom 2015 og 2021 stadfesta at institusjonen ønskjer å ta på seg ansvaret for samlingsporteføljen som eit nasjonalt oppdrag. Dette er eit strategisk val, men eit annleis val enn det Universitetet i Oslo gjorde i 2014. Universitetet i Bergen har signalisert at ord-bokarbeid og utvikling av språkleg infrastruktur for norsk språk skal ha ei vitskapleg plattform og derfor er legitime arbeidsoppgåver for eit universitet. Resultatet kan lesast både i statsbudsjettproposisjonen for 2021, i meldingsdelen i proposisjonen om språklov (Prop 108 L) og i Språkstatus 2021. Universitetet i Bergen framstår der som det nasjonale ord-bokmiljøet i Noreg, og med opprettinga av Termportalen frå 2021 er ansvarsfeltet altså utvida. Dei felte som då er tydeleg definert i dette nasjonale oppdraget, er felt der staten ved Kulturdepartementet har uløyste oppgåver. Det er altså enno ikkje eit “vitskapleg senter for språkleg dokumentasjon” lik det Språkrådet gjorde framlegg om i 2014. Fagleg kan ein nok sjå konturane av det, men framleis ikkje organisatorisk. Og dette “senteret” er altså ikkje ‘uavhengig’ på den måten Språkrådet rådde til.

Oppsummert: Flyttinga av samlingane har gjeve Universitetet i Bergen ny fagleg aktivitet og nye oppgåver, og det er allmenn velvilje og positiv haldning til det. Så er det nok dei i fagmiljøet som kan kjenna seg skuffa over at stadnamndisiplinen ikkje er reetablert og målføregranskninga skubba noko i bakgrunnen. Det skulle vera tydeleg at årsakene til dette ikkje berre ligg i flyttinga av språksamlingane, men like mykje i interne fagpolitiske og økonomiske prosessar ved institusjonen.

## 10 Sluttord – så langt

Oppbygginga av dei nasjonale språksamlingane har ei lang historie – hundre år om ein t.d. tek utgangspunkt i Norsk stadnamnarkiv. I denne artikkelen er det hovudsakleg berre tre år av denne lange perioden som er skildra. Dei som la samlingane til rette ved Universitetsbiblioteket i åra etter 2015, kunne ha fargelagt denne saga endå meir. Dei som stod i prosessane ved Universitetet i Oslo i åra fram til 2014, må vel òg før eller seinare få gje si framstilling av ei intrikat sak. Ei samla og utdjupa fram-

stilling av norsk språksamlingspolitikk krev i alle høve eit større format enn tidsskriftartikkelen.

I den intense prosessen rundt flyttinga av samlingane i 2014 og 2015 var spørsmålet om nasjonalt samarbeid og nasjonalt eigarskap eit viktig tema. Dette perspektivet bør ikkje gløymast. I den digitale tidsalderen er avstanden mellom fagmiljø og samlingar prinsipielt likeverdig same kva fagmiljø og universitet ein snakkar om. Å gjera Språksamlingane ved Universitetet i Bergen til ein nasjonal forskingsressurs kan realiserast ved å invitera fagmiljøa rundt i landet inn i dette arbeidet. Og ikkje minst: ved at fagmiljøa rundt i landet sjølve inviterer seg inn.

## Medieoppslag

På Høyden 4.8.14 = Vil ha språksamlingar til Bergen. Intervju med m.a. Helge Sandøy. <https://pahoyden.khrono.no/sprak-spraksamlingane-uo/vil-ha-spraksamlingar-til-bergen/355513>. Sett 21.9.21.

På Høyden 28.8.14 = Vi er ikkje ute på spekkhoggeri. Intervju med Johan Myking, <https://pahoyden.khrono.no/johan-myking-lle-spraksamlingane/vi-er-ikkje-ute-pa-spekkhoggeri/355794>. Sett 21.9.21.

På Høyden 28.11.14 = Vil ha språksamlingane til Bergen. Intervju med Dag Rune Olsen, Oddrun Samdal, Johan Myking, <https://pahoyden.khrono.no/lle-spraksamlingane/vil-ha-spraksamlingane-til-bergen/357412>. Sett 21.9.21.

På Høyden 10.4.15 = Meiner pengane må følgja med på lasset. Intervju med statssekretær Bjørgulv Vinje Borgundvaag, <https://pahoyden.khrono.no/arkivklynge-lle-spraksamlingane/meiner-pengane-ma-folgja-med-pa-lasset/359636>. Sett 21.9.21.

På Høyden 29.4.15 = Reknar med språkavklaring om nokre månader. Intervju med statssekretær Bjørn Haugstad. <https://pahoyden.khrono.no/kunnskapsdepartementet-spraksamlingane/reknar-med-sprakavklaring-om-nokre-manader/359926>. Sett 21.9.21.

På Høyden 11.11.15 = Få svar frå ministeren. <https://pahoyden.khrono.no/institutt-for-lingvistiske-litteraere-og-estetiske-studium-norsk-ordbok/fa-svar-fra-ministeren/363307>. Sett 21.9.21.

På Høyden 19.4.16 = Emeritusane er klare til dugnad. Intervju med Oddrun Grønvik, Ole-Jørgen Johannessen og Tom Schmidt. <https://pa>

- hoyden.khrono.no/norsk-ordbok-spraksamlingane/emeritusane-erklare-til-dugnad/366113. Sett 21.9.21.
- På Høyden 23.5.19 = Målføresamlingar i uføre. Intervju med m.a. Gjert Kristoffersen, Randi Neteland og Peder Gammeltoft. <https://pa-hoyden.khrono.no/anders-fagerjord-gjert-kristoffersen-johan-myking/malfoeresamlingar-i-ufore/282602>. Sett 21.9.21.
- Uniforum 22.8.14 = Flytt heller språksamlingane til Universitetet i Bergen! Leiarartikkel, <https://www.uniforum.uio.no/meiner/2014/flytt-spraksamlingane-til-universitetet-i-bergen.html>. Sett 21.9.21.
- Uniforum 1.9.14 = Språksamlingane treng ein spesialsydd finansieringsmodell. Reportasje frå møte på Litteraturhuset 28.8.14. <https://www.uniforum.uio.no/nyheter/2014/08/spraksamlingane-ma-fa-ei-skreddarsydd-finansiering.html>. Sett 21.8.21.

## Notat, saksdokument og brev

- Brev LLE 29.04.20 = *LLEs språksamlinger – noen momenter fra fagmiljøet i nordisk språk*. Brev frå fagmiljøet i nordisk til LLE/HF/UiB. UiB/LLE.
- ILN møtenr 3/2014 V-SAK 4 = *Vidare forvaltning og drift av samlingane knytte til leksikografi og namnegransking*. <https://www.hf.uio.no/iln/om/organisasjon/styret/moter/2013-2016/protokoll-2014/protokoll-10-6-2014.html>. Sett 14.9.21.
- LLE/ILN 10.2.15 = *Språksamlingane ved Universitetet i Oslo og eventuell overføring til Universitetet i Bergen*. Brev frå LLE til ILN 10.2.15. UiB/UiO.
- NK 1 = *Språksamlingane og Norsk Ordbok – ny driftsorganisering*. Brev frå Nynorsk kultursentrum til KUD 27.11.14.
- NK 2 = *Språksamlingane og Norsk Ordbok – skisse til ny driftsorganisasjon, Notat 27.11.14*. Vedlegg til NK 1.
- NK 3 = *Prioritert løysing for språksamlingane og Norsk Ordbok*. Brev frå Nynorsk kultursentrum. Underskrive av styreleiar Lodve Solholm. 29.1.15.
- Notat til rektor 26.05.15 = UiB og språksamlingane – organisatoriske moment. UiB/JM.

- UiB 71/2014 = *Protokoll fra møte i universitetsstyret 28.8.2014*, sak 71 <https://www.uib.no/ledelsen/81331/protokoll-fra-møte-i-universitetsstyret-2882014>, sett 14.9.21.
- UiBHF Fsak 46/15 = *Protokoll fra møte i fakultetsstyret ved Det humanistiske fakultet 20.10.2015*, sak 46 [https://www.uib.no/sites/w3.uib.no/files/attachments/protokoll\\_20.10.15.pdf](https://www.uib.no/sites/w3.uib.no/files/attachments/protokoll_20.10.15.pdf). Sett 14.9.21.
- UiO/KD Status = *Status – overtakelse og videre forvaltning av Universitetet i Oslos språksamlinger*. Brev til KD 23.03.15. <https://www.hf.uio.no/iln/om/organisasjon/styret/moter/2013-2016/innkalling-2015/o-sak-3-7-videre-drift-og-forvaltning-av-samlingene.pdf>. Sett 14.9.21.

## Litteratur

- Akselberg, Gunnstein. 2014. Kritisk for norske namn. Lesarinnlegg, *På Høyden* 31.7.14. <https://pahoyden.khrono.no/debatt-sprak/kritisk-for-norske-namn/390115>, sett 21.9.21
- Bang, Anne. 2014. Vår digitale felles hukommelse – et strategisk viktig område for UiB. Lesarinnlegg, *På Høyden* 28.11.2014. <https://pahoyden.khrono.no/debatt/var-digitale-felles-hukommelse-et-strategisk-viktig-omrade-for-uib/390189>, sett 21.9.21.
- Bokmålsordboka og nynorskordboka*. <https://www.ordbokene.no>.
- Helleland, Botolv. 1997. Stadnamna og Bergen. *Kronikk, Bergens Tidende* 12.08.1997. S. 22).
- Kyrkjebø, Rune og Johan Myking. 2017. Språk på flyttefot – Språksamlingane til Bergen. *Årbok for Universitetsmuseet i Bergen*. 91–94.
- Norsk Ordbok a–h (NO-AH)*. <https://www.uib.no/le/138863/norsk-ordbok--h-no-ah>. Sett 15.9.21.
- Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet*. [www.no2014.uib.no](http://www.no2014.uib.no). Sett 17.9.21.
- Norsk ordbokpolitikk. Notat fra Språkrådet til Kulturdepartementet 3. mars 2014*. <https://www.sprakradet.no/upload/Aktuelt/Notat%20om%20norsk%20ordbokpolitikk.pdf>. Sett 26.9.21.
- Prop. 108 L (2019–2020) Lov om språk (språklova)*. <https://www.regjeringen.no/no/dokumenter/prop.-108-l-20192020/id2701451/?ch=1>. Sett 15.9.21.
- Revisjonsprosjektet*. <https://www.uib.no/le/121561/revisjonsprosjektet-bokmålsordboka-og-nynorskordboka>. Sett 15.9.21.



- Solholm, Lodve. 2015. Flytt språksamlingane til Bergen! Kronikk, *Bergens Tidende* 25.2.15. <https://www.bt.no/btmeneringer/kronikk/i/e55Ea/flytt-spraaksamlingane-til-bergen>. Sett 14.1.21.
- [Språksamlingane på Arkivportalen.] *Arkivportalen*. <https://www.arkivportalen.no/search?repository=SPR>. Sett 17.9.21.
- Språkstatus 2021. Språkpolitisk tilstandsrapport frå Språkrådet*. Oslo: Språkrådet. <https://www.sprakradet.no/globalassets/vi-og-vart/publikasjoner/sprakstatus/sprakstatus-2021.pdf>. Sett 15.9.21.
- Tildelingsbrev 2016. *Statsbudsjettet for 2016 kap. 260 – Tildelingsbrev for Universitetet i Bergen*. <https://www.regjeringen.no/contentassets/d5cc2f23a57040b792abo6100f57fofo/tildelingsbrev-uib-2016--1997743.pdf>. Sett 14.6.21.
- Vikør, Lars S. 2016. Språkdokumentasjon og språkforskning. To sider av same sak, eller to saker? I Karlsen, H. m.fl. (red.) 2016. *Livet er øve, og evig er ordet. Festskrift til Norsk Ordbok 1930–2016*. Oslo: Samlaget. 105–131.
- Vikør, Lars S. 2018. *Inn i Norsk Ordbok. Brukarrettleiing og dokumentasjon*. Oslo: Samlaget. <http://no2014.uib.no/eNo/tekst/Inn%20i%20Norsk%20Ordbok.pdf>. Sett 15.9.21.
- Vikør, Lars S. og Åse Wetås. 2016. Norsk Ordbok: om folket – av folket – for folket. Historia om ordboksverket fram til i dag. I Karlsen, H. m.fl. (red.) 2016. *Livet er øve, og evig er ordet. Festskrift til Norsk Ordbok 1930–2016*. Oslo: Samlaget. 15–39.
- Ådland, Einar. 1997. Ein krig i krigen. Striden om Norsk Stadnamnarkiv i Bergen. Bergen: *Årbok for Bergen museum*. 54–60.

## Abstract

In 2014 the University of Oslo decided not to continue its responsibility for the national language collections, thus provoking an extensive public debate. In 2015 the Ministry of Education and Research decided to transfer the collections to the University of Bergen followed by a permanent funding of 6 mill NOK per year. The article describes how the University of Bergen defined its interest in this transfer as well as the various strategic considerations behind this decision. Attempts at rescuing the collections were also made by a national dictionary group representing Norwegian universities. The University of Bergen accepted

JOHAN MYKING

the national responsibility for the collections, which are now in a process of merging with other language collections. Lexicography now forms a major field of activity, and revision work on the standard dictionaries for bokmål and nynorsk as well as Norsk Ordbok receives additional government funding. Development of the other scientific subfields is hampered by the modest permanent funding. The overall scientific goal is to further develop the Language Collections as a national research infrastructure.

Johan Myking  
Universitetet i Bergen  
Institutt for lingvistiske, litterære og estetiske studium  
Boks 7805  
NO-5020 Bergen  
johan.myking@uib.no



# Kollokasjoner – språkets R2-D2 og C-3PO

Av Gunn Inger Lyse, Margunn Rauset og Helge Dyvik

I revisjonen av *Bokmålsordboka* og *Nynorskordboka* treng leksikografane reiskapar som effektivt og etterretteleg finn ordsekvensar som seier noko om typisk bruk og kva frekvente uttrykk ord inngår i. Artikkelen tek føre seg korleis ein definerer og studerer kollokasjonar i ulike tradisjonar, då fleire tilnærmingar til omgrep som *kollokasjon* og *fleirordsuttrykk* møtest i eit praktisk felt som leksikografi. Hovudvekta i artikkelen ligg på korleis ulike reiskapar kan nyttast for å finne kollokasjonar som er relevante å ha med i ordbøkene. Korpuskel-leks er eit korpusgrensesnitt utvikla for leksikografisk bruk. Der kan brukarane søkje etter kollokasjonar ved hjelp av statistiske assosiasjonsmål og ved regulære uttrykk på tvers av mange korpus. NorGramBank er ein trebank med søkjetemplat utvikla for leksikografisk bruk. Der er kollokasjonar ikkje berre eit reint statistisk fenomen av ord som førekjem saman og kan søkjast etter lineært, men søket kan òg referere til syntaktisk struktur og syntaktiske relasjonar mellom orda.

## 1. Innleiing<sup>1</sup>

Ordet *collocation* er belagt i Oxford English Dictionary frå 1750 og har vore brukt om ord som førekjem saman sidan 1917 (Rosqvist 2014: 7). Mange knyter likevel nemninga til den engelske lingvisten J. R. Firth, som på 1950-talet kom med dei bevinga orda «You shall know a word by the company it keeps» (Firth 1957: 11). Det handlar om at vi i vanleg kommunikasjon set saman fleire ord til uttrykk eller heile setningar, slik at

1. Korpuskel-leks og INESS-trebanken NorGramBank er utvikla ved UiB og tilgjengelege via språkinfrastrukturen CLARINO gjennom CLARINO-senteret i Bergen (De Smedt et al. 2015, Rauset et al. 2022).

dei i fellesskap dannar eit tydingsinnhald som igjen skal gje ei viss mein- ing. Tydinga til eitt ord høyrer saman med tydinga til eit anna ord. Sitatet må ikkje tolkast som at tydinga til eit ord einast trer fram når ein kombin-erer det med andre ord, eller at det leksikografiske prosjektet med å definere enkeltord som isolerte einingar er håplaut. Som kompetente språkbrukarar har vi jo som regel ei meining om kva eit ord betyr. Snarare peiker sitatet til Firth på at når vi lærer ord og tydingar i eit språk, gjer vi det ut frå erfaring med korleis dei einskilde orda blir brukte i bestemte språksituasjonar. Kontekst utgjer såleis ei mogleg kjelde til å finne ut noko om tydingane til ord. Når vi seinare tolkar ei ytring i ein bestemt språkbrukssituasjon, gjer vi ikkje det aleine ut frå ei generell oppfatning av kva kvart einskild ord tyder kvar for seg. Snarare bidreg konteksten til å utdjupe, presisere eller nyansere tydinga til einskilde ord.

Leksikografisk inneber sitatet til Firth to ting: I det leksikografiske arbeidet med å identifisere og skildre tydingane til eit ord er autentiske tekstdøme på bruken av eit ord ei viktig empirisk kjelde til kunnskap for leksikografen. Vidare er gode døme på den typiske bruken av eit ord eit nyttig supplement til ein ordbokdefinisjon, nettopp fordi eit viktig aspekt ved å «kjenne eit ord» er å kjenne til karakteristiske kontekstar og vane- messige ordkombinasjonar som ordet blir brukt i.

Kollokasjonsstudiar har blitt eit stort forskingsfelt internasjonalt dei siste tiåra, og ifølgje Stefanowitsch (2020: 233f) fordeler studiane seg i tre hovudgrupper. Den første typen er ei stor mengd studiar som med utforskande metode identifiserer kollokasjonane i store digitale tekst- samlingar, det vi omtalar som korpus. Stefanowitsch hevdar at forskarane bak oftast fokuserer på metode – korleis ein kan preprosessere korpusa, kva assosiasjonsmål ein skal bruke osv. – og i mindre grad på reint ling- vistiske forskingsspørsmål eller einskilde kollokasjonar. Ei anna stor gruppe kollokasjonsstudiar er den bruksretta forskinga som nyttar fre- kvensstudiar til å finne kollokasjonar som er relevante å gjere greie for i ordbøker og anna undervisningsmateriale. Til slutt finst det ei lita gruppe med ofte deskriptive studiar som undersøker kva ord som støttar eller karakteriserer enkeltord eller eit avgrensa utval av ord. Stefanowitsch (2020: 234) peikar på at det generelt er eit relativt fråvære av teoretisk ambisiøse studiar som plasserer seg i denne gruppa, medan ein i norsk samanheng kan slå fast at vi manglar studiar i alle desse tre gruppene, då litteraturen om norske kollokasjonar framleis er nokså sparsam. For å nemne nokre har Fjeld og Vikør skrive om kollokasjonar og andre ord-

forbindelsar frå eit leksikografisk perspektiv (Fjeld og Vikør 2008, Fjeld 2009). Der er òg eit knippe datalingvistiske studiar av kollokasjonar i eit norsk materiale (Andersen 2011 og 2020, Dyvik, Losnegaard og Rosén 2019, Horvati 2005 og Lyse og Andersen 2012).

Med bakgrunn i leksikografi og utvikling av datalingvistiske verktøy for å studere grammatiske og leksikalske fenomen har forfattarane av denne artikkelen med seg perspektiv frå den andre og bruksretta delen av kollokasjonsforskningsfeltet. Artikkelen har like fullt mange trekk frå den første gruppa av kollokasjonsstudiar med vekt på metode og korpus, då formålet med artikkelen er å gjere greie for dei reiskapane i og utanfor Språksamlingane som er tilgjengelege når leksikografane ved Universitetet i Bergen skal velje ut kollokasjonar til standardordbøkene *Bokmålsordboka* og *Nynorskordboka*. Kunnskap om desse reiskapane håpar vi kan vere til nytte også for eit vidare språkvitskapleg miljø. Med Korpuskelleks og INESS (NorGramBank) har vi fått kraftfulle og supplerande verktøy for å finne norske kollokasjonar, og håpet vårt er at fleire skal få auge opp for forskningsfeltet og ta i bruk reiskapane.

I teoridelen (avsnitt 2) klarlegg vi kva ein forskar på når kollokasjonar er studieobjektet. Korleis kan kollokasjonar delast i ulike undergrupper, og korleis kan dei avgrensast i høve til nærskylde språkfaglege omgrep som fleirordsuttrykk, leksikaliserte uttrykk, idiom eller konstruksjonar med partikkelverb? Vi argumenterer for ein fraseologisk typologi med idiomklyngja, ordspråksklyngja og kollokasjonsklyngja som tre sekkestorleikar – som i staden for å prøve å skilje skarpt mellom kategoriane gjer eit poeng ut av å vise at dei ofte overlappar. Denne delen blir avslutta av ein kort gjennomgang av kollokasjonar i standardordbøkene.

I den metodiske hovuddelen ligg vekta på dei to ressursane Korpuskelleks (avsnitt 3) og INESS-trebanken NorGramBank (avsnitt 4). Vi gjer greie for kva type ressursar dette er og korleis dei kan brukast i kollokasjonssamanheng. I avslutninga (avsnitt 5) samanfattar og samanliknar vi korleis ressursane kan brukast og utfyller kvarandre.

## 2. Kva er ein kollokasjon og kva er han ikkje?

### 2.1 Kollokasjonar i den korpusorienterte og systemorienterte tradisjonen

I leksikografien er det vanleg å skilje mellom to kollokasjonsteoriar eller -retningar: den korpusorienterte og den systemorienterte (Svensén 2004:

208). Meir enn at retningane utgjer motsetnader, ser dei ut til å fokusere på ulike spørsmål.

Den aller enklaste måten å forklare ein kollokasjon på, er at det er to eller fleire ord som ofte førekjem i lag. I den korpusorienterte tradisjonen, som har sitt opphav i J. R. Firth, handlar dette om ord som i autentiske tekstar opptre oftare saman enn det som er statistisk sannsynleg, medan ein del forskarar innanfor den systemorienterte retninga trekkjer fram at det handlar om kva ord vi kan kombinere på ein måte som er naturleg og umarkert: «det er slik vi seier det på norsk» – eller det språket ein undersøker. Oftast er det slik at intuisjonen om kva som er «naturleg» eller «umarkert», passar med det som kan bli kvantifisert statistisk som «vanleg».

Når vi som barn lærer eit språk, skjer det i samspel med omgjevningane. Ved å bli snakka til og høyre språket i bruk, møte det i skrift og sjå analogiar med reglar som gjeld for andre ord, lærer vi til dømes korleis vi byggjer opp ei setning med rett ordstilling, korleis ein bøyer ord og alt anna vi kan omtale som produktive grammatiske reglar. Og ved hjelp av desse produktive reglane kan ein setje saman setningar som «dette er vêt til å bli i godt humør av». Kollokasjonar er noko litt anna, det er ferdige pakkar eller einingar som vi lærer som heilskap, som for eksempel *fint vêt* og *dårleg vêt*. Viss ein prøver å byte ut eit ord i kollokasjonen med eit synonym, gjev det fort ordforbindelsen eit litt underleg preg. Det blir uidiomatisk å skifte ut *halde* i *halde eit foredrag* med *forrette*, då det verbet har ein bruksrestriksjon som knyter det til kyrkjelege handlingar, slik vi kjenner det frå kollokasjonane *forrette ved gudsteneste* eller *forrette ei gravferd*. Like uidiomatisk er det når somme direkteomset frå engelsk og seier *gje eit foredrag*.

Det finst forskingsdesign der kollokasjonar er avgrensa til ord som står rett ved sida av kvarandre, men innanfor den korpusorienterte tradisjonen er det vanleg å tillate ein avstand på opp til fem ord mellom komponentane (Stefanowitch 2020: 220). På norsk kan vi gjerne kalle denne avstanden *kollokasjonsrekkevidde* (eng. 'collocation span'). Vi kan illustrere det med ordsekvensen *miste fatninga*, som ein i korpus kan finne i modifisere former som «mista ikkje lett fatninga», «mistar ein augneblink fatninga», «miste dei fullstendig grepet og fatninga», men der det uansett er *miste + fatninga* som utgjer eininga. Ein kan tenkje på orda som inngår i kollokasjonen, som nokre velkjende figurar frå Star Wars-universet, R2-D2 og C-3PO (fig. 1): «R2-D2 er én av to skikkelser som



Figur 1: R2-D2 og C-3PO.  
Foto: Mulyadi på Unsplash

er med i alle filmene i serien. [Den andre] er «protokoldroiden» C-3PO som R2-D2 *stadig opptrer sammen med* (Wikipedia, s.v. «R2-D2,» lesen 18.04.2022, vår kursivering). Analogien til kollokasjonar er at det ofte er andre som er saman med og mellom dei, men det er likevel dei som har den tette koplinga mellom seg i form av syntaktiske relasjonar, ikkje berre lineære posisjonar.

I denne artikkelen, der hovudvekta ligg på verktøya nytta til kollokasjonsanalyse, baserer vi oss mest på den kvantitative korpusorienterte retninga. Kjensla av at to eller fleire ord «heng saman» som ei eining, heng ofte saman med ei erfaring med at dei ofte opptrer saman, og dette er noko vi kan kvantifisere, observere og telje statistisk. Som statistisk omgrep kan vi definere ein kollokasjon som ein statistisk signifikant samførekomst mellom ord (Sag et al. 2002), eller som det er formulert i *Nordisk leksikografisk ordbok*: «Termen kollokasjon kan også defineres ut fra ordenes typiske opptreden i tekstsammenheng, der frekvens og syntaktisk nærhet i konkrete tekster blir lagt til grunn når enkelte kombinasjoner blir betraktet som kollokasjoner» (NLO 1997: 155). Ein slik definisjon er utgangspunktet for statistiske mål på kollokasjon, som vi kjem tilbake til i del 3.2.2. Det å ta utgangspunkt i faktisk språkbruk og mønster ein kan oppdage i korpus, er heilt sentralt i korpuslingvistikken, der kollokasjonsstudiar er eit dominerande felt.

Vi skal likevel seie litt om den systemorienterte retninga, som har vore premissleverandør for mykje av den leksikografiske kollokasjonsforskninga i Norden (jf. t.d. Malmgren 2003, Svensén 2004, Rosqvist 2010 og 2014). Her legg forskarane gjerne vekt på kva ord vi kan kombinere på ein måte som er naturleg og umarkert, og dei har eit restriktivt syn på kva som er å rekne som kollokasjonar. Sentralt står å utvikle kriterium som gjev grupper av samanliknbare einingar med same semantiske relasjon mellom komponentane, som såleis kan gje innsikt i språkssystemet. Den systemorienterte tradisjonen legg til grunn at kollokasjonar har ein hierarkisk struktur med eit hovudord (*basis*) som utgjer den tematiske kjernen, og eit biord (*kollokator*) som støttar eller karakteriserer basis. Basis utgjer den konstante delen av kollokasjonen, medan

kollokatoren til dels kan skiftast ut, utan at det endrar på tydinga til kollokasjonen. Basis blir vanlegvis brukt med ei av dei mest kjende tydingane sine, medan kollokatoren, som blir vald i lag med – ikkje uavhengig av – basisen, ofte er brukt i overført tyding, som verba i *vekkje mistanke* og *sløkkje tørsten* (Svensén 2004, Rosqvist 2014). Fleire stiller krav om at både basis og kollokator må høyre til dei opne ordklassene. Funksjonsord er ikkje sett på som moglege kollokatorar i den systemorienterte tradisjonen, og ordpar som *hus og heim*, *koppar og kar* er då ikkje å rekne som kollokasjonar, då det hadde vore ein ordforbindelse av to likestilte basisar.

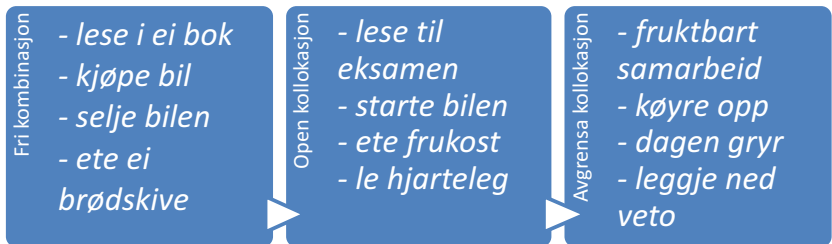
## 2.2 Kriterium for å skilje kollokasjonar frå andre ordforbindelsar

Eit kriterium for å skilje kollokasjonane frå andre ordforbindelsar er fikseringsgrad. I *Nordisk leksikografisk ordbok* blir kollokasjonar rekna til dei mindre faste ordforbindelsane (NLO 1997: 119). Der blir kollokasjon definert som «ordforbindelse der betydningen har en klar sammenheng med betydningen til de enkelte delene» (154) – her blir altså komposisjonalitet trekt fram som det fremste kjenneteiknet ved frasetypen. På linje med den systemorienterte tradisjonen framhevar NLO at eit karakteristisk trekk ved kollokasjonane er at forbindelsen består av ein overordna og konstant basis og ein kollokator med karakteriserande funksjon.

Eit anna kriterium er variasjon – å sjå på kor fritt utskiftbare komponentane i frasen er. Når forskarane innan den systemorienterte tradisjonen gjer eit poeng av å drage linjer mellom kategoriane *frie kombinasjonar*, *kollokasjonar* og *idiom*, sjølv om det kan vere vanskeleg, er det til dels motivert av eit behov for å identifisere dei ordkombinasjonane som bør ordbokførast. Kollokasjonar ser dei som ein slags halvfabrikat ein kan hente fram frå minnet når ein treng dei, og som dei meiner ein bør gjere greie for i ordbøker, medan dei frie kombinasjonane, som blir kombinerte på føreseieleg vis, ikkje høyrer heime i ordbøker (Svensén 2004: 210). Leksikografane Fjeld og Vikør (2008) har til liks med *Nordisk leksikografisk ordbok* (1997) ei litt anna kategorisering, der dei reknar med to typar av kollokasjonar. Fig. 2 byggjer på skiljet i desse kjeldene mellom frie kombinasjonar, opne kollokasjonar og avgrensa kollokasjonar, som ei form for nyansering av den systemorienterte tradisjonen – og med ei presisering om at det ikkje finst faste grenselinjer mellom dei (Fjeld og Vikør 2008: 112). Frå venstre i figuren står dei *frie*



*kombinasjonane*, dei vi produserer ved hjelp av dei produktive syntaktiske reglane i eit språk. Dei fleste kombinasjonar av ord er frie, og det betyr at det enkelte ordet ikkje har avgjerande styring eller innverknad på valet av dei andre orda, men vi kan skifte dei ut alt etter kva ord vi treng for å uttrykkje det vi ønskjer. Kollokasjonar er underlagt større kombinatoriske restriksjonar, men likevel i ulik grad. Difor skil figuren mellom *opne kollokasjonar* (der kollokatoren til ei viss grad kan skiftast ut, utan at det endrar på tydinga til eininga (NLO 1997: 125)) og *avgrensa kollokasjonar* (som høyrer til dei faste ordforbindelsane som ikkje tillèt endring i kollokatoren utan at det endrar tydinga til eininga (NLO 1997: 117)).



Figur 2: Frie kombinasjonar og opne og avgrensa kollokasjonar.

Ser vi på den frie kombinasjonen «lese i ei bok», kan ein ifølgje dei semantiske og syntaktiske reglane i norsk like gjerne kjøpe ei bok eller selje henne, slå opp i eller leggje henne frå seg osv. *Lese til eksamen* har eit sterkare band mellom orda for å skildre prosessen med å lære seg pensum før ein eksamen. Når vi likevel har karakterisert kollokasjonen som open, er det fordi ein kan erstatte verbet med ein del andre verb frå same semantiske domene som *pugge til eksamen* og *studere til eksamen*. At vi har plassert *le hjarteleg* som ein open kollokasjon, handlar om at det finst eit knippe utskiftbare adverb som *inderleg*, *høgt*, *godt* osv., som i liten grad endrar på tydinga til eininga.

Dei avgrensa kollokasjonane heilt til høgre er dei som i minst grad har utskiftbare komponentar, og vi kan omtale dei som dei sterkaste kollokasjonane. Eit fellestrekk ved ein del av dei er at dei inneheld komponentar som sjeldan blir kombinert med andre enn dei vi ser her. Til dømes kan *leggje ned veto* ikkje uttrykkjast på så mange andre måtar.

Det er ikkje semje om at eit partikkelverb som *køyre opp* (for å *ta lappen* eller *få sertifikat*) høyrer heime i ein slik modell, ettersom den systemorienterte retninga legg til grunn at funksjonsord ikkje er moglege kollokatorar. Ein viktig grunn til å inkludere partikkelverb i slike modeller er at ein bør gjere greie for ordforbindelsen i ordbøker. Tek vi t.d. *vaske opp* 'reingjere koppar og kar' og *vaske ned* 'ta storreingjering av hus', illustrerer dei godt at mange av partikkelverba er ikkje-komposisjonelle og leksikaliserte uttrykk. Når dei t.d. kan stå utan utfylling, som i «eg køyrd opp i stad» og «i morgon skal eg vaske opp», har dei klare trekk av å vere kollokasjonar, men grensedraginga mot idiom er krevjande. Eitt viktig skilje mellom dei to typane ordforbindelsar er at vi vanlegvis brukar kollokasjonen som heilskap ikkje-metaforisk, sjølv om kollokatoren gjerne blir brukt i overført tyding, t.d. i *gripe sjansen* (Svensén 2004: 211).

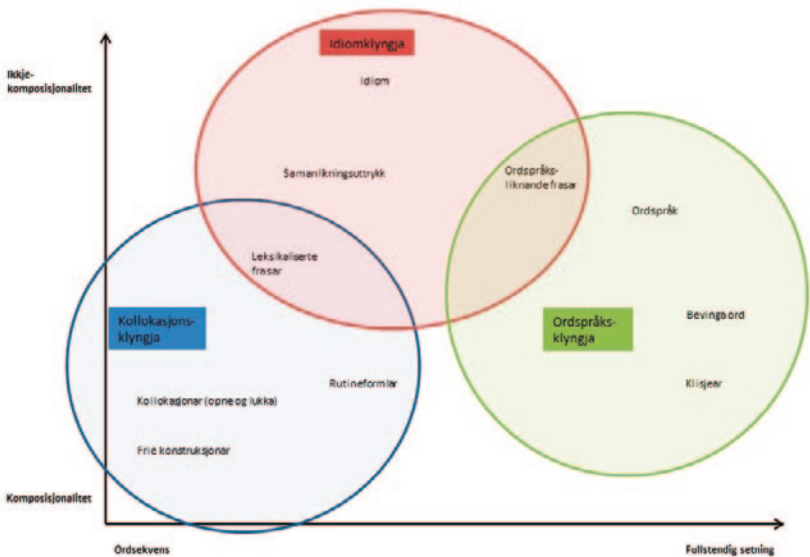
Fjeld og Vikør (2008: 113f) er dessutan blant dei som skil mellom grammatiske og leksikalske kollokasjonar. Dei grammatiske kollokasjonane har funksjonsord, ikkje minst preposisjonar, som biord, og ser vi på døme som *kunnskap om noko* og *erfaring med noko*, blir det tydeleg at ein må treffe med riktig preposisjon for å uttrykkje seg idiomatisk. Det er eit argument for å ha slike med i ordbøker som ein konvensjonalisert måte å uttrykkje eit innhald på. Den systemorienterte Svensén (2004: 213) hevdar i *Handbok i lexicografi* at denne typen ordforbindelsar ikkje er å rekne som kollokasjonar, men konstruksjonar.

Dei leksikalske kollokasjonane er kjenneteikna av at både hovudordet og biordet har sjølvstendige tydingar, som *hoppe strikk*, *hoppe bukk*, *gullande rein* og *sove søtt*. Elles ville *ta lappen* og *få sertifikat*, som vi nemnde ovanfor, begge med verb og substantiv som innhaldsord, vere rekna som leksikalske kollokasjonar etter systemorienterte kriterium.

I denne artikkelen baserer vi oss på den fraseologiske typologien som er utarbeidd i Rauset (2022: 37 ff), og som er oppsummert i fig. 3. Figuren gjer greie for eit utval sentrale frasetypar ut frå to dimensjonar: om det er ein ordsekvens eller ei fullstendig setning (x-aksen), og komposisjonaltitet (y-aksen). Idiomklyngja og ordspråksklyngja representerer dei faste frasane (av typen ordspråk som *morgonstund har gull i munn*, bevinga ord som *Nu hvil dig borger, det er fortjent*, klisjear som *fordi du fortjener det*, ordspråksliknande frasar som *først til mølla*, idiom som *ha mange jern i elden*, samanlikningsuttrykk som *som olja lyn* og leksikaliserte frasar som *kald krig*), medan kollokasjonane er plassert innanfor sirkelen nedst til venstre, som del av dei mindre faste frasane i kollokasjonsklyngja. Figu-

ren skil mellom opne og avgrensa kollokasjonar, men ikkje eksplisitt mellom grammatiske og leksikalske (i tråd med argumentasjonen i denne artikkelen om at det er behov for å gjere greie for grammatiske kollokasjonar på linje med leksikalske i ordbøker). Typologien er basert på det kjende fraseologiske kontinuumet fri kombinasjon – kollokasjon – idiom/faste frasar (sjå t.d. Barkema 1996: 125, og som vi kjenner frå den systemorienterte tradisjonen), men tek sikte på å gjere greie for eit større utval av fleirordsuttrykk.

I fig. 3 blir det framheva at kollokasjonar er meir komposisjonelle enn idiom, samanlikningsuttrykk og andre faste frasar i dei romslege idiom- og ordspråksklyngjene, utan at ein hevdar at ein fullt ut kan slutte seg til tydinga til eininga ut frå tydinga til enkeltkomponentane i kollokasjonen (jf. partikkelverba *vaske opp* og *vaske ned*). Som før nemnt er eit problem ved dette kriteriet at det er gråsoner mellom kva som er fullt komposisjonelt og ikkje. I mange forskingsprosjekt vil det vere nødvendig med mykje klarare kriterium og grensedragingar, men bak figuren ligg ei praktisk-leksikografisk erkjenning av at det er svært krevjande, men ofte mindre nødvendig i leksikografisk samanheng, å skulle skilje dei ulike frasetypane frå kvarandre, og at dei ofte kan overlappe.



Figur 3: Typologi som skil kollokasjonsklyngja frå dei faste frasane i idiom- og ordspråksklyngjene.

### 2.3 Kollokasjonar i Bokmålsordboka og Nynorskordboka

I revisjonen av *Bokmålsordboka* og *Nynorskordboka* skal vi oppdatere innhaldet i to ordbøker med lang trykt og digital historie (Rauset 2019, Revisjonsprosjektet 2022). Leksikografane har tre hovudoppgåver: sikre at definisjonane er i tråd med dagens språkbruk, peike ut eit moderne og relevant ordtilfang i begge skriftspråka, og gjere utvalet av lemma i ordbøkene likare (Selback 2020). Leksikografane treng empirisk materiale for å vurdere lemmatilfang, fleirordsuttrykk, ordtydingar og syntaktisk åtferd, og finne bruksdøme. Det er då avgjerande med gode søkjeverktøy som lét leksikografane jobbe etterretteleg og samstundes effektivt (Lyse 2020).

Standardordbøkene gjer ikkje greie for kollokasjonar (i den strenge tydinga av ordet, til skilnad frå t.d. idiom) som ein eigen kategori, men dei blir vanlegvis nytta som bruksdøme. Om tydinga til eininga avvik så mykje frå tydinga til komponentane at ho treng forklaring, legg vi kollokasjonane blant dei faste uttrykka. Dersom ein hadde laga heilt nye ordbøker, hadde det i dag vore naturleg å skilje ut kollokasjonar som ein eigen kategori, slik det t.d. er gjort i *Islex*, der kollokasjonane er gjorde greie for under dei ulike tydingane til ordet. I den ordboka kjem dei faste uttrykka (idiomklyngja og ordspråksklyngja i fig. 3) til slutt i ordartiklane, som signal om at avstanden mellom dei inngåande orda og den regulære tydinga i dei faste uttrykka er større enn i kollokasjonane.

## 3. Kollokasjonsstudiar i Korpuskel-leks

I arbeidet med kollokasjonar i standardordbøkene har vi fleire ressursar, og i denne artikkelen gjer vi greie for to av dei. Det eine verktøyet er Korpuskel-leks, som lét oss gjere avanserte søk på enkeltord og frasar (Lyse 2020). Det andre verktøyet er trebanken NorGramBank for bokmål og nynorsk, der kvar setning har ein detaljert syntaktisk analyse (Dyvik et al. 2016).

### 3.1 Korpuskel-leks

Korpuskel er eit grensesnitt for å tilgjengeleggjere, søkje i og analysere tekstkorpus, og er eit generisk verktøy med over 60 korpus på ulike språk, format og mediatypar (Meurer 2012).<sup>2</sup> Korpuskel-leks er ei vidare-

2. <https://clarino.uib.no/lex-prod/corpus/>

utvikling av Korpuskel. Dette spesialtilpassa grensesnittet for leksikografisk bruk vart implementert for Revisjonsprosjektet i 2018, og lèt brukaren søkje på tvers av mange korpus samstundes. Som del av CLARINO er Korpuskel-leks ope tilgjengeleg, innanfor rammene av brukslisensane som gjeld for dei enkelte tekstkorpusa.

I tab. 1 listar vi dei 13 korpusa som per i dag ligg i Korpuskel-leks. Kvart av dei er lista med kolonnar som viser skriftspråk, storleik i millionar løpeord og skiljeteikn, tidsrommet korpuset dekkjer, og ei grovkategorisering av sjanger og tilgang. Metadata med meir informasjon om kvart korpus finn ein via nettsidene til Korpuskel-leks. Dette er eit heterogent materiale som er samla inn til ulike formål, og det kan difor ikkje kallast eit balansert korpus.

Styrken i materialet i Korpuskel-leks er mengda. Samla gjev Korpuskel-leks i dag eit materiale på 3,1 milliardar teikn, altså ord og teiknsetjing. Det totale materialet for bokmål er på 2,9 milliardar teikn, medan det totale materialet på nynorsk er på dryge 217 millionar, altså med ei fordeling på 93 % for bokmål og 7 % for nynorsk. Materialet har tekstar frå perioden 1765–2021, men med ei hovudvekt på nyare materiale. Om lag 75 % av materialet er tekst frå 1998 og fram til notid, altså om lag 2,2 milliardar teikn. Dette er naturleg nok viktig for revideringa av standardordbøkene, som skal beskrive eit moderne ordtilfang. Ser vi på skriftspråka kvar for seg, er 77 % av bokmålsmaterialet (2,9 milliardar) frå 1998 og framover i

Korpusnamn	Skriftspråk	Storleik (mill.)	Tidsrom	Sjanger	Tilgang
Forskning.no (2017)	nob	26,2	1998–2017	tekst: sakprosa	avgrensa
NAOB-tekster	nob	238,7	1846–2018	tekst: skjønnlitteratur	avgrensa
NBs frie tekster (bokmål)	nob	516,4	1765–2013	tekst: blanda	open
NBs frie tekster (nynorsk)	nno	46,0	1850–2010	tekst: blanda	open
Norsk avis korpus (bokmål)	nob	2046,3	1998–2021	tekst: sakprosa	open
Norsk avis korpus (nynorsk)	nno	21,1	1998–2020	tekst: sakprosa	open
Norsk avis korpus annotert	nob	35,7	2001–2009	tekst: sakprosa	open
Norsk ordboks nynorsk korpus	nno	129,1	1866–2012	tekst: blanda	avgrensa
Norsk ordboks nynorsk korpus (2017)	nno	4,0	1926–2017	tekst: blanda	avgrensa
Talk Of Norway	nob, nno	64,3	1998–2016	tekst: sakprosa	open
Dialektending (Talebank)	nno	5,6	1960–2020	tale: dialekt korpus	avgrensa
Industristad (Talebank)	nno	2,3	1948–2013	tale: dialekt korpus	avgrensa
Talesøk (Talebank)	nno	1,9	1998–2016	tale: dialekt korpus	avgrensa

Tabell 1: Korpus som er søkbare gjennom Korpuskel-leks.

tid; tilsvarende tal for nynorsk materialet (217 millionar) er 56 %. Det må understrekast at tilgangen på eldre tekst òg er nyttig i ordbokarbeidet, mellom anna for å vurdere tradisjonsord og i kva grad dei er i bruk i dag samanlikna med tidlegare.

Ein veikskap i materialet er den skeive fordelinga mellom nynorsk og bokmål, og dessutan sjangerfordelinga. Tab. 2 viser at sakprosa utgjer nesten 70 % av det samla materialet, gjennom avistekstar og populærvitenskapleg tekst (*Norsk aviskorpus bokmål og nynorsk, Forskning.no 2017* og *Talk of Norway*). I tillegg er der noko sakprosa i materialet som ligg i kategorien «blanda», og som både inneheld skjønnlitterær tekst og sakprosa (*Nasjonalt bibliotekets (NBs) frie tekstar bokmål og nynorsk* og dei to nynorsk korpuser).

Sjanger	Storleik (mill.)	Prosent
skjønnlitteratur	238,74	7.61 %
sakprosa	2193,61	69.91 %
blanda (skjønnlitteratur, sakprosa)	695,52	22.17 %
dialektkorpus	9,78	0.31 %

Tabell 2: Sjangerfordeling mellom korpuser i tabell 1.

### 3.2 Korpusstudiar i Korpuskel-leks

Vi vil her gjennomgå dei to sentrale metodane i korpusstudiar; enkle søk og frekvenslister, og statistisk baserte rangeringar av ordsekvensar med høg kollokasjonsstyrke. Dataa i alle døma i del 3.2 kjem frå Korpuskel-leks, med søk i alle korpuser i tab. 1, bortsett frå den annoterte delen av Aviskorpuset. For kvart søk er søkje-URL gjeve opp i ein fotnote. URL-en gjev direkte det same søket, og føreset at brukaren har tilgang til dei aktuelle korpuser.

To terminologiske presiseringar må til. For det første omtalar vi for det meste sekvensar av ord. Men i eit elektronisk korpus finn vi teknisk sett strengar, òg kalla *teikn* (engelsk ‘tokens’) skilde med mellomrom, som omfattar både ord og teiknsetjing. Det vil seie at vi òg kan møte sekvensar av ord og teiknsetjing.

For det andre omtaler studiar med assosiasjonsmål vanlegvis ordsekvensar som *n*-gram, altså ein sekvens av ord med ei lengd på *n*. Eit *n*-gram med ei lengd på 2 kallar vi til dømes eit bigram, medan eit trigram har ei lengd på 3. Assosiasjonsmåla i Korpuskel-leks er implementerte som ei utrekning på forbindelsen mellom to teikn for å identifisere dei

komponentane som stadig opptrer saman, òg sjølv om dei ikkje alltid følger rett etter kvarandre – jamfør analogien til R2-D2 og C-3PO i avsnitt 2.1. Vi må difor skilje mellom bigram og trigram osv., som altså er ordsekvensar med direkte kontakt mellom orda, og statistisk sterke ordforbindelsar, som ikkje føreset slik direkte kontakt. I Korpuskel kan ein be om ei rangering av kollokasjonar for eit ord, t.d. *fatning*, med eit kontekstvindauge på t.d. [ $\pm 2$ ]. Då vil det bli rekna ut kollokasjonsstyrke for *fatning* og alle ord som kjem i høvesvis første og andre posisjon framfor og etter. Eit setningsdøme som: *han tek tapet med fatning og seier at (...)* ville då gje følgjande ordpar som vi kan kalle kollokasjonskandidatar, skrivne mellom hakeparentesar, og der ein understrek indikerer at orda ikkje står direkte etter kvarandre: [tapet\_ fatning], [med fatning], [fatning og] og [fatning \_seier].

### 3.2.1 Regulære uttrykk og enkle frekvenslister

Ein kan gjere enkle søk på ordsekvensar ved hjelp av regulære uttrykk. La oss bruke verbet *hoppe* som døme. Eit typisk behov for leksikografen er å undersøkje kva ordsekvensar dette ordet inngår i. I Korpuskel-leks søker ein etter ein streng ved å skrive han mellom hermeteikn, og vi kan søkje etter ein vilkårleg streng ved å skrive ein tom hakeparentes. Søk (1) under vil vere eit søk på ordforma «hoppe» og det første ordet som følgjer etterpå (eit vanleg bigram), medan søk (2) representerer det tilsvarende trigrammet. Søk (3) vil gje treff på trigram som har ordforma «hoppe» i midten. Ofte er det nyttig å snevre inn eit søk tidsmessig, t.d. berre til treff etter 1970, som i søk (4). Dette er til hjelp dersom ein treng å undersøkje moderne bruk. Ofte treng vi å sjå alle bøyingsformene (søk 5), og i nokre tilfelle søker vi på leksemet «hoppe» (det vil seie alle bøyingsformene), som i søk (6). Det siste gjer redaksjonen relativt sjeldan, fordi det avgrensar treffa til det subsettet av korpus som har morfologisk analyse tilgjengeleg – dvs. korpusa *Norsk aviskorpus (annotert)*, *Dialektkorpus*, *Forskning.no (2017)*, *Industristad*, *Nynorskkorpusa*, *Talesøk* og *Talk of Norway*. Sjå elles nettsidene til Korpuskel-leks for ein meir omfattande dokumentasjon av søkjeuttrykk.<sup>3</sup>

(1) “hoppe” []

3. <https://clarino.uib.no/lex-prod/documentation/korpuskel-documentation>

- (2) “hoppe” [][]<sup>4</sup>
- (3) [] “hoppe” []<sup>5</sup>
- (4) “hoppe” [][] :: year > “1970”<sup>6</sup>
- (5) “hoppe|hopper|hoppa|hoppet|hoppa” [][]<sup>7</sup>
- (6) [/hoppe/ & {verb}] [][]<sup>8</sup>

Søket i døme (2) gjev 45478 treff, fordelte på 12 korpus og innanfor tidsrommet 1838–2021.<sup>9</sup> Korpuskel-leks viser treffa på ulike måtar. Fana *Konkordans* viser treffa til søket i det klassiske visingsformatet med ei linje per treff (KWIC, eller KeyWord In Context). Treffordet/-orda er då midtstilte med ei viss mengd kontekst før og etter. Fana *Ordliste* oppsummerer kor mange gonger kvart mønster førekom, og er enkel å laste ned, som i uttrekket i tab. 3. Denne tabellen viser toppen av ei frekvensliste over korleis treffa på søket "hoppe" [][] er fordelte på dei ulike mønstera, med sekvensen «hoppe ut av» som den mest frekvente sekvensen.<sup>10</sup>

I det daglege ordbokarbeidet er slike frekvenslister eit effektivt verktøy for å få eit raskt overblikk over ordsekvensar som kan seie noko om

4. Søkje-URL: <https://clarino.uib.no/lex-prod/corpus/avis-plain,avis-nno,dialekt,fn-new,industristad,naob,nb-fri-nob,nb-fri-nno,nnk,nnk-new,talesoek,ton/%22hoppe%22%20%5B%5D%5B%5D>
5. Søkje-URL: <https://clarino.uib.no/lex-prod/corpus/avis-plain,avis-nno,dialekt,forskning-no,fn-new,industristad,naob,nb-fri-nob,nb-fri-nno,nnk-new,talesoek,ton/%22hoppe%22%20%5B%5D%5B%5D>
6. Søkje-URL: <https://clarino.uib.no/lex-prod/corpus/avis-plain,avis-nno,dialekt,fn-new,industristad,naob,nb-fri-nob,nb-fri-nno,nnk,nnk-new,talesoek,ton/%22hoppe%22%20%5B%5D%5B%5D%20%3A%3A%20year%20%3E%20%221970%22%20>
7. Søkje-URL: <https://clarino.uib.no/lex-prod/corpus/avis-plain,avis-nno,dialekt,fn-new,industristad,naob,nb-fri-nob,nb-fri-nno,nnk,nnk-new,talesoek,ton/%22hoppe%7Chopper%7Choppa%7Choppet%7Choppa%22%20%5B%5D%5B%5D>
8. Søkje-URL: <https://clarino.uib.no/lex-prod/corpus/avis-plain,avis-nno,dialekt,fn-new,industristad,naob,nb-fri-nob,nb-fri-nno,nnk,nnk-new,talesoek,ton/%5B%E2%88%A5hoppe%E2%88%A5%20%26%20%7Bverb%7D%5D%20%5B%5D%5B%5D%20>
9. Legg merke til at vi ikkje søker i det annoterte Aviskorpuset i Tabell 1 samstundes som vi søker i dei to vanlege avisorpusa på høvesvis bokmål og nynorsk, ettersom dette ville kunne gje dupliserte treff.
10. Ein feil i ein del av inndata til Korpuskel-Leks førte til feilaktige treff. Denne feilen kjem til å bli fiksa i korpuset; i denne omgangen nytta vi eit søkjeuttrykk som ser bort frå den delen av korpus der feilen ligg. I staden for søkjeuttrykket i døme (2) på s. 48, nytta vi difor dette uttrykket: "hoppe" [][] @@ 0-93552955, 96452303-10000000000.



typisk bruk, og peike oss til frekvente uttrykk der fokusordet vårt inngår. I lista i tab. 3 ser vi t.d. sekvensar som *hoppe etter Wirkola*, *hoppe i det*, *hoppe i havet* og *hoppe i taket*. I søk på trigram, som dette dømet, kan vi òg intuitivt kjenne att toordssekvensar som er moglege kollokasjonskandidatar, som *hoppe av* og *hoppe bukk*. Slike reine frekvenslister er dessutan nyttige for å danne seg eit bilete av vanlege uttrykksmåtar som kan vere relevante som illustrerande bruksdøme i ordbokartikkelen (*hoppe ut frå*, *hoppe i fallskjerm*).

frekvens	"hoppe" [ ] [ ]	frekvens	"hoppe" [ ] [ ]
982	hoppe ut av	304	hoppe bukk over
571	hoppe etter Wirkola	299	hoppe opp og
563	hoppe i det	291	hoppe i taket
515	hoppe i sjøen	256	hoppe i havet
502	hoppe på ski	238	hoppe av ,
456	hoppe inn i	233	hoppe av isen
447	hoppe ut i	232	hoppe over til
374	hoppe av .	231	hoppe finaleomgangen .
326	hoppe i fallskjerm	229	hoppe i vannet
314	hoppe ut fra	213	hoppe av og

Tabell 3: Toppen av frekvenslista over treordsekvensar med «hoppe» som første ordform ved søk i Korpuskel-leks, med søk i alle korpus i Tabell 1 unnateke den annoterte versjonen av Aviskorpuset.

Frekvenslister som denne fangar berre inn dei ordsekvensane som samla har *høgast frekvens*. Men dei tek ikkje omsyn til i kva grad orda som inngår i ein sekvens, statistisk sett har sterkare tendens til å opptre saman enn med andre ord – altså om dei er ein sterk kollokasjon. Til dette har ein det ein kallar assosiasjonsmål.

### 3.2.2 Assosiasjonsmål

Såkalla assosiasjonsmål (AM) er statistiske metodar som prøver å kvantifisere ein intuisjon om ord som «hører saman». Enkelt sagt går dei fleste AM ut på å analysere relasjonen mellom kor ofte orda i ein sekvens opptrer saman, samanlikna med kor ofte dei førekjem kvar for seg.

I Korpuskel-leks er fem ulike mål på kollokasjon implementerte per i dag: *Frekvens*, *Relativ frekvens*, *Log Likelihood (LL)*, *Mutual Information (MI)* og *MI \* log (frekvens)*. Vi fokuserer her på korleis slike statistiske AM-mål kjem til nytte i det leksikografiske arbeidet, og for den interesserte lesaren viser vi til dokumentasjonen på nettsidene til Korpuskel-

leks og til Evert (2004), Lyse og Andersen (2012) og Pedersen et al. (2011).

I tab. 3 ovanfor såg vi ei frekvenssortert liste over treordssekvensar som startar med hoppe, altså eit søk på “hoppe” [][]. Fig. 4 viser tilsvarende dei ordpara som blir rangert høgast på statistisk grunnlag når *hoppe* er første ord og ein ser på kva ord som førekjem i posisjon ein eller to etter *hoppe*. I Korpuskel-leks kan vi søkje etter statistisk sterke kollokasjonar under fana *Kollokasjonar* (tredje fane øvst på skjermbiletet i fig. 4). Her kan ein velje mellom ulike attributt som er knytte til søkjeordet, og som regel vel vi den bøygde ordforma (attributtet *word*). Nedtrekksmenyane *venstrekontekst* og *høgrekontekst* handlar om kollokasjonsrekkjevidde og spesifiserer kor mange ord på kvar side ein vil ha for å berekne kollokasjonskandidatane til søkjeordet, og ein kan velje ein verdi frå 0 (ingen kontekst på den eine sida av søkjeordet) til 6 (alle kontekstord opp til 6 ord unna søkjeordet). For å svare til søket som gav frekvenslista i tab. 3, vel vi *høgrekontekst* = 2, altså to «plassar» til høgre for søkjeordet. I feltet *Terskel* i fig. 4 kan ein setje ei nedre grense for kor frekvente kollokasjonskandidatane må vere for å bli vurderte. Det er lurt å ikkje

Value	Frequency	Relative	Log likelihood	Mutual information
hoppe_Wirkola	575	0.15481961	218854.69	13.366307
hoppe_bukk	365	0.10951095	226106.02	12.866794
hoppe_fallskjerm	344	0.062352728	173673.23	12.054243
hoppe_finaleomgangen	299	0.024725048	92299.52	10.719764
hoppe_strikk	106	0.051986266	271264.03	11.791922
hoppe_flyvingen	36	0.13235295	463425.62	13.140109
hoppe_skiflyging	135	0.03736507	213916.58	11.315481
hoppe_sprette	124	0.03856921	225654.72	11.36124
hoppe_høyet	104	0.043315284	254868.72	11.528667
hoppe_trampoline	78	0.052419353	301999.62	11.80389

Figur 4: Dei høgast rangerte kollokasjonane for treordsekvensar med «hoppe» som første ordform ved søk i Korpuskel-leks.

setje ei slik grense før ein har sett den første rangeringa utan ein terskelverdi. Ser ein at det kom mange lågfrekvente og mindre interessante ordpar høgt på lista, kan ein eventuelt setje ei grense slik at alle kollokasjonskandidatar med frekvens lågare enn (t.d.) 5 blir filtrerte vekk. I Korpuskel-leks blir alltid alle fem statistiske mål rekna ut for kvart ordpar, og i resultatlista ser ein dei statistiske verdiane i kvar si kolonne. Unntaket er det statistiske målet  $MI * \log(Freq)$ , som berre er produktet av verdiane i dei to siste kolonnane. I nedtrekksmenyen *Sorter etter* vel ein kva for statistisk mål ein vil sortere lista etter. Her viser vi resultatata med målet  $MI * \log(Freq)$ , som ofte viser seg å gje den mest nyttige rangeringa. Det er slik sett kanskje noko uheldig at rekkefølga av kollokasjonskandidatane i figuren ikkje samsvarer med nokon av kolonnane. Grunnen er ganske enkelt at dette femte statistikk målet vart innført i etterkant av at dei fire første var implementerte, medan visinga med fire kolonnar vart ståande.

Samanliknar ein topptreffinga i fig. 4 med den reine frekvenslista som vi såg i tab. 3, ser vi at orda i kvar kollokasjon intuitivt er langt meir bundne til kvarandre enn det som kom på topp i den reine frekvensordlista. Kvar linje (kvart rangerte ordpar) har ei kolonne *Frequency*. Vi ser at dei høgast rangerte ordpara i dette søket er relativt frekvente, som kollokasjonskandidatane [hoppe – Wirkola] (som var funne 575 gonger) og [hoppe bukk] (som var funne 365 gonger). Understrekinga i den første av desse kollokasjonskandidatane betyr at det står eit ord mellom ord1 og ord2. Dette stemmer sjølvsagt òg med intuisjonen om dette dreier seg om uttrykket *hoppe etter Wirkola* 'å prøve seg mot nokon som er betre'. Den statistisk sterke kollokasjonen i dette uttrykket gjeld altså sambandet mellom dei to innhaldsorda, *hoppe* og *Wirkola*, og ikkje mellom *hoppe* og *etter* eller *etter* og *Wirkola*.

Dette er ei statistisk rangering av kollokasjonskandidatar som ikkje tek stilling til lingvistisk analyse. For leksikografen er altså dette eit supplement til vanlege lister av konkordansar eller ordlistefrekvens, men det er ikkje lister som gjev ferdige svar på kva ein eventuelt skal gjere med det som statistisk står som sterke kollokasjonar. Leksikografen vurderer deretter i kva grad dette kvalifiserer til å stå som eit fast uttrykk i ordboka med ei eiga forklaring, om det illustrerer ein bestemt bruk av fokusordet som bør inn som ei eiga tyding, eller om dette er ein illustrerande kollokasjon som kan brukast som bruksdøme under ei eksisterande tyding (uttrykket *hoppe etter Wirkola* er utan tvil eit fast og ikkje-komposisjonelt

idiom som treng eit eige oppslag og ei eiga forklaring i ordbøkene, og ligg allereie som eit eige oppslag der).

Kva så med den kollokasjonen som var rangert som nest sterkast i fig. 4, *hoppe bukk*? Intuitivt fortel språkkjensla oss at ordet *bukk* står ubøygde i denne kollokasjonen – dette er ikkje ein vanleg transitiv relasjon der ein t.d. kan «hoppe bukken». Søkjer ein etter ordsekvensen [] “bukk” i Korpuskel-leks (eit søk etter kva som helst direkte framfor ordforma *bukk*), vil ein raskt finne at verbet *hoppe* kan førekome i ulike bøyingsformer (med høgfrekvente treff som «hoppe bukk», «hopper bukk», «hoppet bukk»).

For å undersøkje kor variabelt uttrykket er, kan ein opne for strengar mellom «hoppe» og «bukk», og òg sjekke kva strengar som kjem imellom, t.d. slik:

“hopp.\*” [][0,3] “bukk” []

Dette søket spesifiserer at ein vil ha alle strengar som begynner på «hopp» (*hoppe, hoppar, hoppa* osv.). Så kan det vere mellom 0 og 3 strengar før strengen «bukk», medan ein tom hakeparentes etter strengen «bukk» betyr at ein vil sjå kva strengar som følgjer rett etter «bukk». Dette gjev ei frekvensliste der ein raskt ser av dei 20 mest frekvente ordsekvensane (tab. 4) at det stabile er ulike bøyde former av verbet *hoppe*, ordforma *bukk*, og for det meste er *bukk* følgt av ordet *over*: *hoppe bukk over*. Så ser ein at det kan skytast inn modifiseringar, som t.d. «hopper statsråden ganske enkelt bukk over» og «hopper også bukk over». På dette grunn-

frekvens	kollokasjon	frekvens	kollokasjon
335	hopper bukk over	10	hoppes bukk over
304	hoppe bukk over	9	hoppe bukk "
202	hoppet bukk over	9	hopper elegant bukk over
30	hoppa bukk over	7	hopper han bukk over
20	hoppar bukk over	7	hopper man bukk over
19	hopper de bukk over	6	hoppe bukk og
16	hopper statsråden ganske enkelt bukk over	6	hoppet bukk i
14	hoppe bukk .	5	hopp bukk over
13	hopper også bukk over	5	hoppe bukk eller
11	hoppe bukk ,	5	hopper helt bukk over

Tabell 4: Toppen av frekvenslista ved søk på variantar i uttrykket : “hopp.\*” [][0,3] “bukk” [] i Korpuskel-leks, med søk i alle korpus i Tabell 1 unntatte den annoterte versjonen av Aviskorpus.

laget er det velmotivert å leggje inn *hoppe bukk over* som eit eige uttrykk med tydinga ‘ikkje ta omsyn til’. I tillegg bør ordparet *hoppe bukk* (om barneleiken) gjerast greie for, sidan heller ikkje dette er ei gjennomsiktig og generell tyding av verbleksemet *hoppe*.

Den tredje høgast rangerte kollokasjonen i fig. 4, [*hoppe* – fall-skjerm], illustrerer ein statistisk sterk kollokasjon som intuitivt ser ut til å vere ganske komposisjonell. Den noverande første tydinga av verbet *hoppe* (etter revisjon) er slik: ‘ta sats og sprette opp i eller gjennom lufta’.<sup>11</sup> Det å ta sats og sleppe seg gjennom lufta i fallskjerm verkar som ei heilt gjennomsiktig tyding, der «hoppe i fallskjerm» eventuelt kunne stå som eit døme på typisk bruk av verbet *hoppe* i tyding 1.

På denne måten kan leksikografen jobbe seg gjennom lister av statistiske kollokasjonar. Men ein må ha i mente at dei ikkje fortel meir enn at her er ein ordsekvens med ei sterk kollokasjonsstyrke. Om dette er ein bit av eit større uttrykk, og kva ein skal gjere med kollokasjonen, er likevel noko leksikografen må vurdere.

Som vi såg av dømet med *hoppe etter Wirkola* i fig. 4, kan det vere at assosiasjonsmåla ikkje fangar inn uttrykk med kollokasjonsrekkevidde på meir enn to ord, sidan dei berre reknar statistikk på ordpar. Ved å leggje inn ein tom streng i søket som dannar utgangspunkt for å rekne ut kollokasjonar, blir det rekna assosiasjonsmål på treordsuttrykk framfor på toordsuttrykk. Eit døme kan vere adjektivet *fruktbar*, som står oppført som del av kollokasjonen *fruktbart samarbeid* i fig. 2 i avsnitt 2.2, og som òg blir brukt som døme på søk i ein syntaktisk trebank i avsnitt 4.2. Med ei søkjeining som

“**fruktba(r|rt)**” []

får ein treff på ordformene *fruktbar* eller *fruktbart* følgt av ein vilkårleg streng (som t.d. «fruktbar dialog» eller «fruktbart virkemiddel»). Ein kan deretter få rangert kva som er statistisk sterke kollokasjonar til desse to-ordsekvensane. Av fig. 5 ser vi at det systemet gjer, er først å hente ut gode toordssekvensar (*fruktbar alder*, *fruktbart samarbeid*, *fruktbart å*) og å vise kva som er statistisk sterke kollokasjonar til desse igjen (*i fruktbar alder*, *et fruktbart samarbeid*, *fruktbart samarbeid med*, *være fruktbart å*). Så må leksikografen framleis vurdere om dette er sekvensar som kan fun-

11. <https://ordbokene.no/nn/31276/hoppe>

Vis kollokasjoner    Attributt: word     ignorer storskriving    Venstrekontekst: 1    Høyrekontekst: 1    Tersket:

Sorter etter: MI \* log(Freq)   

6786 kollokasjoner: 1 2 3 4 5 ... 227    Gå til side:     Last ned

Value	Frequency	Relative	Log likelihood	Mutual information
i fruktbar alder	418	0.000005453039	<input type="checkbox"/>	<input type="checkbox"/>
et fruktbart samarbeid	309	0.00002391944	<input type="checkbox"/>	<input type="checkbox"/>
fruktbart samarbeid med	169	0.0000062779964	<input type="checkbox"/>	<input type="checkbox"/>
være fruktbart å	162	0.00003300429	<input type="checkbox"/>	<input type="checkbox"/>
mer fruktbart å	153	0.000043486973	<input type="checkbox"/>	<input type="checkbox"/>
og fruktbart samarbeid	115	0.0000017304623	<input type="checkbox"/>	<input type="checkbox"/>
fruktbart samarbeid mellom	108	0.00004907969	<input type="checkbox"/>	<input type="checkbox"/>
fruktbar alder .	99	6.042466e-7	<input type="checkbox"/>	<input type="checkbox"/>
lite fruktbar Midtøsten-reise	98	0.00013218306	<input type="checkbox"/>	<input type="checkbox"/>
fruktbar Midtøsten-reise Rapport	95	0.0033389567	<input type="checkbox"/>	<input type="checkbox"/>
lite fruktbart å	92	0.00012409023	<input type="checkbox"/>	<input type="checkbox"/>
en fruktbar måte	88	0.0000028333395	<input type="checkbox"/>	<input type="checkbox"/>
fruktbart samarbeid .	80	4.882801e-7	<input type="checkbox"/>	<input type="checkbox"/>

Figur 5: Dei høgast rangerte kollokasjonane for *fruktbar*/*fruktbart* + *kva* som helst, med éin streng framfor eller éin streng etter. Visning av rangeringa med assosiasjonsmålet  $MI * \log(\text{frekvens})$ .

gere som gode bruksdøme på ei eksisterande tyding av *fruktbar*, eller om det bør leggjast inn som eit fast uttrykk.

Generelt er erfaringa til redaksjonen at det løner seg å starte med regulære søk på frekvente ordsekvensar som fokusordet vårt inngår i, før ein eventuelt òg sjekkar rangeringa med kvart av dei moglege assosiasjonsmåla. Ulike ord har ulik frekvensprofil, og det er lite føreseieleg kva kollokasjonar som kan vise seg nyttige for det aktuelle ordet ein jobbar med. Generelt har Log Likelihood ein tendens til å favorisere høgfrekvente, formulaiske sekvensar som typisk inneheld minst eitt grammatisk ord, som preposisjonar, subjunksjonar eller infinitivsmærket *å*. (t.d. *være fruktbart å* eller *et fruktbart og*). Mutual Information har ein tendens til å favorisere lågfrekvente kollokasjonskandidatar der eitt eller begge orda opptrer særleg eller berre i denne konteksten. Målet  $MI * \log(\text{frekvens})$  gjev større vekt til frekvensen til kollokasjonskandidaten som heilskap, og dette målet er sett opp som standardmålet når ein ikkje har valt noko anna. Dette er ikkje basert på ei vitenskapleg undersøking, men erfaringsmessig gjev dette målet ofte dei mest relevante kollokasjonane med innhaldsord.

Det kan òg nemnast at redaksjonen har tilgang til å bruke ordskissene i Sketch Engine (Kilgarriff et al. 2014). Sketch Engine ligg på ein måte i grenseland mellom det vi finn i Korpuskel-leks og i NorGramBank i INESS. Sketch Engine er eit komplett verktøy for å søkje på ord og få forslag til statistisk sterke kollokasjonar, og ein kan få desse sortert i ulike grammatiske kategoriar (med kategoriar som *modifiers of*, *subjects of*, *objects of*, *and/or*). Denne ressursen er likevel sjeldan den første vi slår opp i, primært fordi det norske tekstmaterialet i Sketch Engine består av nedlasta tekst frå norske nettsider, som dekkjer alt frå offisielle nettsider til bloggtekst. Som diskutert i Lyse (2020: 4) kviler Revisjonsprosjektet på tilgang til kjeldegrunnlag av ein viss skriftleg kvalitet, og vi prioriterer difor empirisk materiale der teksten har vore gjennom ein redaksjonell prosess før publisering (i motsetnad til t.d. ein del bloggtekstar).

#### 4. Kollokasjonsstudiar i INESS

##### 4.1 Trebanken NorGramBank

Ein trebank er eit syntaktisk analysert tekstkorpus – eit korpus der kvar setning er forsynt med ein syntaktisk (nokre gonger òg ein semantisk) analyse. Analyseformata kan variere, og trebankar kan vere bygde opp gjennom reint manuell analyse, gjennom manuell analyse i kombinasjon med større eller mindre innslag av automatisk analyse, eller gjennom reint automatisk analyse (parsing).

NorGramBank (Dyvik et al. 2016) er ein norsk trebank som er utvikla ved INESS-prosjektet ved Universitetet i Bergen (sjå Rosén et al. (2012) og nettsida for prosjektet <http://clarino.uib.no/iness>). Setningane i NorGramBank er analyserte automatisk med den komputasjonelle norske grammatikken NorGram, utvikla gjennom fleire år ved Universitetet i Bergen og basert på den syntaktiske teorien leksikalsk-funksjonell grammatikk (LFG). NorGramBank inneheld no ca. 160 millionar ord analysert tekst (av dette ca. 150 millionar på bokmål), som omfattar aviser, sakprosa, skjønnlitteratur, stortingsforhandlingar og somme andre teksttypar i mindre omfang. Det skjønnlitterære materialet og mykje av sakprosaen har prosjektet motteke i OCR-skanna form frå Nasjonalbiblioteket.

#### 4.2 Nokre søkjetemplat som er eigna i kollokasjonsstudiar

Ein trebank tillèt søk etter og teljing av syntaktiske eigenskapar i tekstar. Dette gjer det mogleg å søkje etter fleirordsuttrykk på grunnlag av syntaktisk struktur og grammatiske relasjonar i uttrykka, og ikkje berre på grunnlag av den lineære plasseringa av elementa i dei. Eit døme kan vere kollokasjonen *fruktbart samarbeid*, som vi har vore inne på tidlegare. I vurderinga av om dette dømet er ein kollokasjon, er det mellom anna relevant å sjå på relative frekvensar. I jamføringa av frekvensane for ulike adjektiv som modifierer substantivet *samarbeid*, er det då nyttig å få med døme der adjektivet ikkje berre står attributivt framfor substantivet, men òg predikativt i ulike konstruksjonar. Då kan lineært søk bli vanskeleg, medan ein trebank kan nyttast. Søkjeuttrykket for å finne adjektiv i ulike syntaktiske funksjonar kan bli komplekst, men til hjelp for leksikografane har NorGramBank ein reiskap kalla *søkjetemplat*, utvikla i samråd med leksikografane i Revisjonsprosjektet og i ordboka NAOB. Eit søkjetemplat er eit ferdig, parametrisert søkjeuttrykk der brukaren berre supplerer parameterverdiar, til dømes lemmaformer det skal søkjast etter. Namnet på templatet for søk etter adjektiv som kan modifisere eit bestemt substantiv, er **N-adjmod(@N)**. Når dette er valt frå ein meny, kjem brukaren til ei side der substantivet kan skrivast inn (fig. 6). Dette søket finn 729 ulike adjektiv, fordelt på 7 303 døme. *fruktbart* kjem på 28. plass med 43 treff; toppen av frekvenslista ser ut som i fig. 7.

**Template:** \* N-adjmod(@N)

**Description:** The adjectives modifying a noun

Lists, with frequencies, all adjectives modifying the noun @N, attributively or predicatively.

**Parameters:**

@N:

Run query

Figur 6: Eit søkjetemplat med parameterverdi supplert av brukaren.

Nytta av å kunne søkje etter predikativ funksjon går til dømes fram ved adjektivet *viktig*, der nokre av dei 105 døma er viste i fig. 8. Desse døma kunne elles berre ha vorte funne gjennom lineært søk med stor kollokasjonsrekkevidde, som ville ha gjeve mykje støy i form av irrelevante treff.



729 match types, 7303 matches. | Page 1 of 15  | Go to page:  Click on a row to see the matching sentences. | Copy format:  plain  NAOB

Count	# noun: atom	# p: value
1071	samarbeid	god
508	samarbeid	nær
478	samarbeid	internasjonal
454	samarbeid	nordisk
406	samarbeid	tett
172	samarbeid	økonomisk
130	samarbeid	interkommunal
127	samarbeid	europpeisk
125	samarbeid	borgerlig
108	samarbeid	forpliktende
105	samarbeid	viktig
105	samarbeid	politisk
104	samarbeid	bred
90	samarbeid	konstruktiv
83	samarbeid	intim

Figur 7: Toppen av frekvenslista for adjektiv som modifierer samarbeid.

Count	# noun: atom	# p: value
125	samarbeid	borgerlig
108	samarbeid	forpliktende
105	samarbeid	viktig

Page 1 of 6		<input type="button" value="Previous"/>	<input type="button" value="Next"/>	Go to page: <input type="text"/>		<input type="button" value="Go"/>	<input type="button" value="Download"/>	
Click on a row to go to the sentence. Mouse over a row to see the structures.								
Treebank	Document	Trans.	Id	Sentence				
nor-stortinget_4	s140527	no	1435	Samarbeid mellom barnevernstjenesten og foreldrene blir ansett som viktig for at hjelpetiltakene skal ha en funksjon.				<input type="button" value="Copy"/>
nor-stortinget_4	s140605	no	1071	Samarbeid med næringslivsaktører i utviklingspolitikken blir derfor viktigere.				<input type="button" value="Copy"/>
nor-stortinget_4	s140619	no	4190	Finanskrisen viste oss at samarbeid på tvers av landegrens er viktig for å løse internasjonale problemer.				<input type="button" value="Copy"/>
nor-stortinget_4	s141105	no	1308	Jeg vil legge til at et styrket samarbeid mellom de enkelte nordiske lands utenrikstjenester faktisk er viktig av flere grunner.				<input type="button" value="Copy"/>
nor-stortinget_4	s141111	no	762	Sjelden er et velfungerende samarbeid på tvers av landegrensene viktigere enn i slike situasjoner.				<input type="button" value="Copy"/>

Figur 8: Døme med samarbeid modifisert av adjektivet viktig.

Trebanken kan òg nyttast til å skaffe oversyn over argumentrammer til verb, og over kva for verb eit substantiv oftast er argument til. Slik informasjon er nyttig i arbeidet med å finne typiske døme på verb-substantiv-kombinasjonar til ordbokopplaga. Vi kan illustrere med eit døme frå fig. 2 i avsnitt 2.2, som viser nokre frie kombinasjonar og opne og avgrensa kollokasjonar. *kjøre bil* og *selje bilen* er nemnde som døme på frie kombinasjonar, medan *starte bilen* er døme på ein open kollokasjon. Ei jamføring mellom desse uttrykka treng informasjon om kor ofte *bil* førekjem som argument ved ulike verb. Til dette kan ein nytte templatet

2234 match types, 26039 matches. | Page 1 of 45  | Go to page:  Click on a row to see the matching sentences. | Copy format:  plain  NAOB

Count	#A-arg2of: value	#B-noun: atom	#C-arg1of: value
1350		bil	være
1248		bil	stå
1081	ha	bil	
1040	kjøre	bil	
939		bil	kjøre
898		bil	komme
580	parkere	bil	
571	være	bil	
445		bil	stanse
363	ta	bil	
348	se	bil	
333	starte	bil	
309		bil	stoppe&stop
274	stanse	bil	
264		bil	svinge
258		bil	kunne
249		bil	passere
249		bil	exist
243		bil	skulle
240	kjøpe	bil	

Figur 9. Verb med bil som argument 2 (patiens) til venstre og som argument 1 (agens) til høyre. (exist symboliserer presenteringskonstruksjon med vere/være, som i: det er få bilar i gatene).

**N-argofverbs(@N)** med parameterverdien *bil*. Dette søket får 26 039 treff fordelt på 2 234 verb-argument-kombinasjonar; dei 20 mest frekvente kombinasjonane er viste i fig. 9.

*selge* kjem på 51. plass med 85 treff. (Som tidlegare nemnt byggjer ikkje skiljet mellom frie kombinasjonar og kollokasjonar berre på skilnaden mellom frekvensane til uttrykka – ein må òg ta omsyn til frekvensane til kvart ord, som ein òg kan finne i trebanken.)

At trebanken ikkje berre tillèt søk etter syntaktiske funksjonar som subjekt og objekt, men òg etter argumentposisjonar, som svarar til semantiske roller, gjer at ei vidare mengd døme blir funne. Det gjeld til dømes setningar der *bil* som argument 2 (patiens) ikkje er objekt, men subjekt til passiv eller hovud for eit partisipp; sjå fire av dei 85 døma på *selge bil* i fig. 10, der berre det siste har *bil* som objekt til aktiv *selge*.

Grammatikken NorGram, som har gjeve opphav til dei syntaktiske analysane i NorGramBank, inneheld òg informasjon om mange fleirordsuttrykk, til dømes grammatiske kollokasjonar med selekterte preposisjonar eller partiklar, og visse idiom. Ein kan søkje etter døme på slike grammatikk-analyserte fleirordsuttrykk med templat som **V-mwe(@V)**

Nye biler som selges i dag, har en helt annen sikkerhet enn gamle biler.

Det betyr at det selges flere biler med lavere avgift, som igjen betyr at flere betaler avgift.

I klimaforliket fra juni i 2012 ble det enstemmig definert - med Fremskrittspartiets subsidiære tilslutning - et tak i 2017, eller 50 000 solgte biler.

Det er altså fire årstall som de som produserer drivstoff, importerer biler og selger biler, forholder seg til.

Figur 10. Ulike døme på bil som argument 2 til selge.

**Template:** \* V-mwe(@V)

**Description:** Multi-word expressions with a verb

Lists, with frequencies, types of multi-word expressions (MWEs) with the verb @V, typically expressions with selected prepositions or selected particles, or verb phrase idioms with selected content words.

**Parameters:**

@V:

Figur 11. Templatet V-mwe(@V) med gripe som parameterverdi.

Count	#p: atom
3920	gripe*inn
1817	gripe*fatt*i
1806	gripe*etter
1618	gripe*om
1349	gripe*til
1033	gripe#tak*i
304	gripe*an
11	gripe*fatt*om

Figur 12. Søkjeresultat: grammatikk-analyserte fleirordsuttrykk med gripe som hovudord.

og N-mwe(@N), som finn fleirordsuttrykk med eit bestemt verb eller substantiv som hovudord. Fig. 11 viser templatet V-mwe(@V) med verbet *gripe* som parameterverdi. Dette søket gjev treffa i fig. 12.

I slike høve blir ikkje frekvens eit kriterium for brukaren for å identifisere fleirordsuttrykk, sidan dei allereie er identifiserte i NorGram. Men det er ikkje uvanleg at grammatikken har funne meir enn ein analyse av slike konstruksjonar, sidan mange fleirordsuttrykk med ikkje-komposisjonelt innhald i tillegg vil ha ein gyldig komposisjonell analyse. *Han tenkjer på hytta* kan både tyde at han har hytta i tankane (grammatisk kollokasjon med selektert preposisjon), og at han tenkjer når han er på hytta (fri kombinasjon med adverbial preposisjonsfrase). Analysen i trebanken vil då

ofte vere vald på statistisk grunnlag. Dette inneber at søk etter slike grammatikk-analyserte fleirordsuttrykk ofte bør supplerast med søk etter tilsvarende frie kombinasjonar med komposisjonelt innhald, både fordi statistikken kan ha valt feil i nokre høve, og fordi det kan finnast fleirordsuttrykk som ikkje er dekte av grammatikken. Det finst òg templat for søk etter slike frie kombinasjonar.

I avsnitt 2 er den systemorienterte retninga nemnd, der kollokasjonar blir analyserte med eit hovudord og ein kollokator. I døme som *le hjertelig*, *nekte blankt*, *tvile sterkt* og *protestere heftig* er verba hovudord. For at ein skal kunne identifisere slike, er det viktig å kunne jamføre med alternative moglege kollokatorar ved dei same verba og sjå på frekvensar

Count	#verb: atom	#w: value
1273	le	høyt
289	le	lavt
284	le	hjertelig
214	le	godt
143	le	hånlig
138	le	kort
123	le	rått
94	le	nervøst
88	le	hysterisk
84	le*av	høyt

Figur 13. Toppen av søkjeresultatet: dei mest frekvente adjektiva som står adverbialt til verbet *le*.

og tydingar. Sidan trebanksøk kan referere til ordklassar og syntaktiske relasjonar, blir det mogleg å finne alle slike kandidatar gjennom eit enkelt søk etter ord i den relevante syntaktiske funksjonen – ein treng ikkje å søkje etter alle moglege kollokatorar enkeltvis. Templatet for å finne adjektiv som står adverbialt til eit bestemt verb, er **V-adverbialadj(@V)**. Eit søk med verbet *le* som parameterverdi gjev 6016 treff fordelt på 808 adjektiv; fig. 13 viser toppen av frekvenslista, der *hjertelig* kjem på tredje plass.

## 5. Oppsummering

Utgangspunktet for denne artikkelen har vore korleis ein definerer og studerer kollokasjonar i ulike tradisjonar, og korleis dei to verktøya Korpuskel-leks og NorGramBank kan nyttast i fraseologiske studiar av kollokasjonar, særleg med tanke på dei som kan vere relevante å ta med i ordbøker.

Både i den korpusbaserte tradisjonen og den systemorienterte tradisjonen ser forskarane på kollokasjonar som to eller fleire ord som «heng saman» som ei leksikalsk eining, og at kollokasjonar er ein slag friare forbindelse mellom ord enn t.d. idiom, der skilnaden frå meir frie til meir faste uttrykk heng saman med frekvens, med om tydinga til eininga er

komposisjonell, og med om ein kan variere delar av uttrykket utan å endre tydinga.

Noko av det som skil retningane, er at den systemorienterte tradisjonen har eit meir restriktivt syn på kva som er å rekne som kollokasjonar enn den korpusbaserte. I den systemorienterte er det sentralt å utvikle kriterium for å finne grupper av samanliknbare einingar med same semantiske relasjon mellom komponentane for å få innsikt i språkssystemet. I denne artikkelen argumenterer vi snarare for ein fraseologisk typologi med ei idiomklyngje, ei ordspråksklyngje og ei kollokasjonsklyngje som tre sekkestorleikar, der vi i staden for å prøve å skilje skarpt mellom kategoriane gjer eit poeng av å vise korleis dei ofte overlappar.

Den korpusbaserte tradisjonen oppfattar kollokasjonar som noko som er uavhengig av syntaktisk analyse, og som først og fremst dreier seg om den lineære plasseringa av ord i høve til kvarandre og at to (eller fleire) ord førekjem saman oftare enn ein skulle vente ut frå frekvensane deira kvar for seg. Dette er metoden som ligg bak bruken av Korpuskelleks, der vi har vist døme på å søkje etter ordsekvensar med hjelp av regulære uttrykk eller gjennom statistiske assosiasjonsmål. Regulære uttrykk gjer det enkelt å finne dei mest frekvente mønstera av ord, medan assosiasjonsmål kan løfte fram samførekomstar som ikkje nødvendigvis er høgfrekvente, men som likevel er statistisk sterke kollokasjonar. Denne tilnærminga er enkel og effektiv, men seier berre at visse ord har ein tendens til å førekomme saman. Det blir leksikografen sin jobb å vurdere om dette er ordsekvensar som står som typiske døme på ei bestemt tyding av eit ord, eller om kollokasjonen representerer ei eiga tyding som bør få sitt eige oppslag i ordboka. Dette er altså ein bruk av omgrepet *kollokasjon* som er langt meir open enn den vi finn i den systemorienterte tradisjonen.

Til samanlikning lèt NorGramBank oss søkje etter grammatiske mønster som er knytte til eit søkjeord, gjennom søk på syntaktisk struktur og grammatiske relasjonar. Dette opnar for å kunne finne syntaktiske variantar av same kollokasjon, som til dømes predikativ eller attributiv plassering av adjektiv, og det gjer det mogleg å søkje etter alle ord i ein viss syntaktisk relasjon til eit hovudord som eit grunnlag for å skilje ut dei som bør analyserast som kollokatorar (som i dømet *le hjarteleg*). NorGramBank inneheld dessutan informasjon om mange fleirordsuttrykk, til dømes visse idiom, og grammatiske kollokasjonar med selekterte preposisjonar eller partiklar, som ein då kan søkje direkte etter.

Slike trebank-søk kan såleis gje eit inventar av kollokasjonskandidatar med ord som ein så kan studere nærare i det mykje større Korpuskel-korpuset. Frå ståstaden til leksikografen utfyller dermed desse to ressursane kvarandre.

## Litteratur

### Ordbøker

*Bokmålsordboka og Nynorskordboka*. Språkrådet og Universitetet i Bergen. <ordbokene.no> (april 2022).

*Islex*. Árni Magnússon-instituttet for islandske studiar, Det Danske Sprog- og Litteraturselskab, Universitetet i Bergen, Göteborgs universitet, Fróðskaparsetur Føroya og Helsingfors universitet. <islex.no> (april 2022)

NAOB = Det Norske Akademis ordbok. <naob.no> (april 2022).

NLO = *Nordisk leksikografisk ordbok*. Bergenholtz, Henning, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jon Hilmar Jonsson og Bo Svensén (red.). 1997. Skrifter utgitt av Nordisk forening for leksikografi. Universitetsforlaget.

### Annan litteratur

Andersen, Gisle. 2011. Evaluation of alternative association measures for extraction of terminology based on a large Norwegian corpus. *SYNAPS – A Journal of Professional Communication* 26/2011, 62–68.

Andersen, Gisle. 2020. Phraseology in a cross-linguistic perspective: A diachronic and corpus-based account. *Corpus Linguistics and Linguistic Theory* 18(2), 365–389. <https://doi.org/10.1515/cllt-2019-0057>.

Barkema, Henk. 1996. Idiomaticity and terminology: A multi-dimensional descriptive model. *Studia Linguistica* 50 (2), 125–160.

De Jong, Franciska, Bente Maegaard, Koenraad De Smedt, Darja Fišer & Dieter Van Uytvanck. 2018. CLARIN: Towards FAIR and Responsible Data Science Using Language Resources. I: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*, Miyazaki, Japan. European Language Resources Association (ELRA), 3259–3264. <https://aclanthology.org/L18-1515.pdf>

De Smedt, Koenraad, Gunn Inger Lyse Samdal, Rune Kyrkjebø, Hemed Ali Al Ruwehy, Øyvind Liland Gjesdal, Victoria Rosén & Paul Meurer. 2015. The CLARINO Bergen Centre: Development and

- Deployment. I: Selected Papers from the CLARIN Annual Conference 2015, October 14–16, 2015, Wrocław, Poland, 1–12. Linköping Electronic Conference Proceedings 123:1. <http://www.ep.liu.se/ecp/123/001/ecp15123001.pdf>
- Dyvik, Helge, Gyri Smørðal Losnegaard & Victoria Rosén. 2019. Multiword expressions in an LFG grammar for Norwegian. I: *Representation and parsing of multiword expressions: Current trends*, red. Yannick Parmentier & Jakub Waszczuk, 69–108. Berlin: Language Science Press. <https://doi.org/10.5281/zenodo.2579037>
- Dyvik, Helge, Paul Meurer, Victoria Rosén, Koenraad De Smedt, Petter Haugereid, Gyri Smørðal Losnegaard, Gunn Inger Lyse & Martha Thunes. 2016. NorGramBank: A ‘Deep’ Treebank for Norwegian. I: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, red. Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk & Stelios Piperidis, 3555–3562. Portorož, Slovenia: ELRA.
- Firth, John Rupert. 1957. A synopsis of linguistic theory 1930-1955 (1957). I: *Studies in Linguistic Analysis*, ei spesialutgåve av Philological Society, red. J.R. Firth, 1–32, Oxford: Blackwell.
- Fjeld, Ruth Vatvedt. 2009. Leksikografisk dokumentasjon av flerordsenheter i norsk. *LexicoNordica* 16: 103–118.
- Fjeld, Ruth Vatvedt og Lars S. Vikør. 2008. *Ord og ordbøker: Ei innføring i leksikologi og leksikografi*. Kristiansand: Høyskoleforlaget.
- Evert, Stefan. 2004. The Statistics of Word Co-occurrences: Word Pairs and Collocations. Ph.D. thesis, IMS, University of Stuttgart.
- Horvati, Eszter. 2005. Automatisk gjenkjenning av norske kollokasjoner. Masteroppgåve, Universitetet i Oslo
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. 2014. The Sketch Engine: ten years on. *Lexicography* 1(1), 7–36. Springer.
- Lyse, Gunn Inger & Gisle Andersen. 2012. Collocations and statistical analysis of n-grams. I: *Exploring Newspaper Language: Using the Web to Create and Investigate a large corpus of modern Norwegian*, red. Gisle Andersen, 79–109. Studies in Corpus Linguistics 49. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.49.05lys>

- Lyse, Gunn Inger. 2020. Ut med *adamsslekt* og inn med *arveprinsesse*? Leksikografiske metodar i revisjonen av *Bokmålsordboka* og *Nynorskordboka*. I: *Nordiska studier i lexikografi* 15, 215–224.
- Malmgren, Sven-Göran. 2003. Begå eller ta självmord? Om svenska kollokationer och deras förändringsbenägenhet. *Språk och Stil* 13 (ny följd), 123–168.
- Meurer, Paul. 2012. Corpuscle – a new corpus management platform for annotated corpora. I: *Exploring Newspaper Language: Using the Web to Create and Investigate a large corpus of modern Norwegian*, red. Gisle Andersen, 29–50. *Studies in Corpus Linguistics* 49. Amsterdam: John Benjamins. <https://doi.org/10.1075/scl.49.02meu>
- Pedersen, Ted, Satanjeev Banerjee, Bridget McInnes, Saiyam Kohli, Mahesh Joshi & Ying Liu. 2011. The Ngram statistics package (text::nsp): A flexible tool for identifying ngrams, collocations, and word associations. I: *Proceedings of the workshop on multiword expressions: From parsing and generation to the real world*, red. Paul Cook et al., 131–133.
- Rauset, Margunn. 2019. Bokmålsordboka og Nynorskordboka – einnegga, toegga eller siamesiske tvillinger? *LexicoNordica* 26: 155–175.
- Rauset, Margunn. 2022. *Frasar til besvær? Studiar av norm og bruk i norsk fraseologi*. Doktorgradsavhandling, Universitetet i Bergen. <https://hdl.handle.net/11250/2992943>
- Rauset, Margunn, Gyri Smørdal Losnegaard, Helge Dyvik, Paul Meurer, Rune Kyrkjebø & Koenraad De Smedt. 2022. Words, words! Resources and tools for lexicography at the CLARINO Bergen Centre. I: *CLARIN. The infrastructure for language resources*, red. Darja Fišer og Andreas Witt, 537–560. Berlin: deGruyter.
- Revisjonsprosjektet. 2022. Revisjonen av Bokmålsordboka og Nynorskordboka. Oppdatert 11. mars 2022. <https://revisjonsprosjektet.no>.
- Rosén, Victoria, Koenraad De Smedt, Paul Meurer & Helge Dyvik. An open infrastructure for advanced treebanking. I: Jan Hajič, Koenraad De Smedt, Marko Tadić & António Branco (red.) *META-RESEARCH Workshop on Advanced Treebanking at LREC2012*, 22–29, Istanbul, Turkey, May 2012.
- Rosqvist, Bodil. 2010. Ordförbindelser i SAOB – En undersøkning av beskrivningen av kollokationer. *Nordiska Studier i Lexikografi* 10, 455–469.



- Rosqvist, Bodil. 2014. Hårt arbete och sträng vila. I: *Svenska kollokationer i lexicografisk och lexikologisk belysning*. Göteborgs universitet. <http://hdl.handle.net/2077/35259>
- Sag, Ivan, Timothy Baldwin, Francis Bond, Ann Copestake & Ann Flickinger. 2002. Multiword Expressions: A Pain in the Neck for NLP. I: *Computational Linguistics and Intelligent Text Processing. CICLing 2002. Lecture Notes in Computer Science 2276*, red. Gelbukh, A. Berlin og Heidelberg: Springer, 1–15. [https://doi.org/10.1007/3-540-45715-1\\_1](https://doi.org/10.1007/3-540-45715-1_1)
- Selback, Bente. 2020. «Å nei, det ordet er ikkje lov på nynorsk!» Eller ...? Om parallell redigering av to norske ordbøker. I: *Nordiska studier i lexicografi 15*, 297–305.
- Stefanowitsch, Anatol. 2020. *Corpus linguistics: A guide to the methodology*. Textbooks in Language Sciences 7. Berlin: Language Science Press. <https://doi.org/10.5281/zenodo.3735821>
- Svensén, Bo. 2004. *Handbok i lexicografi : Ordböcker och ordboksarbeite i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.
- Unsplash. Bilete av R2-D2 og C-3PO. Foto: Mulyadi. Sett 18.04.22. [unsplash.com/photos/JEfw\\_d\\_okQGE](https://unsplash.com/photos/JEfw_d_okQGE)
- Wikipedia, s.v. «R2-D2,» lesen 18.04.2022, <https://no.wikipedia.org/wiki/R2-D2>.

## Abstract

In the revision of the Norwegian dictionaries *Bokmålsordboka* and *Nynorskordboka*, the lexicographer needs efficient and user-friendly tools that can point to typical uses of individual words, and frequent phrases associated with them. This article presents and discusses the definition and study of collocations in different traditions, as multiple approaches to terms such as ‘collocation’ and ‘multiword expression’ meet in the practical field of lexicography. The main part of the article presents available digital tools for finding and studying collocations that may be candidates for inclusion in a dictionary. Corpuscle-lex is a corpus search tool adapted for lexicographical use. The user may search for collocations using statistical association measures and regular expressions across multiple corpora. NorGramBank is a treebank with syntactically analysed text. Hence a search may refer to syntactic structures and grammatical rela-

GUNN INGER LYSE, MARGUNN RAUSET OG HELGE DYVIK

tions and not only to the linear positions of words in a text. NorGram-Bank comes with search templates enabling the lexicographer to make complex searches for the grammatical patterns around a word without engaging directly with query language expressions.

Gunn Inger Lyse  
Universitetet i Bergen  
Institutt for lingvistiske, litterære  
og estetiske studium  
Postboks 7805  
NO-5020 Bergen  
gunn.lyse@uib.no

Margunn Rauset  
Universitetet i Bergen  
Institutt for lingvistiske, litterære  
og estetiske studium  
Postboks 7805  
NO-5020 Bergen  
margunn.rauset@uib.no

Helge Dyvik  
Universitetet i Bergen  
Institutt for lingvistiske, litterære  
og estetiske studium  
Postboks 7805  
NO-5020 Bergen  
helge.dyvik@uib.no



## Med heimstaddialekten i bagasjen?

### Om korleis målføresynopsisen og folketeljningane kan brukast i historisk sosiolingvistik

Av Randi Neteland

Artikkelen tek utgangspunkt i ei teoretisk problemstilling som gjeld faktorene som spelar inn i koinéformingsprosessar, nemleg om det er ein direkte samanheng mellom dialektbakgrunnen til innflyttarane til ein ny stad og språktrekka som blir brukte i den nye dialekten som etterkvart veks fram. Dette spørsmålet blir testa med å kombinere folketeljninga frå den nye byen Narvik i 1900 og målføresynopsisen sine presise data om tradisjonelle dialektar i Noreg. Medan mange tidlegare forskingsprosjekt om talemål i industristader og liknande nye byar har konkludert med at innflyttarane sine talemål har vesentleg innverknad på sluttproduktet av koinéformingsprosessen, viser denne undersøkinga at dialektane som innflyttarane hadde med seg i bagasjen frå heimlassen sin, ikkje kan ha hatt direkte innverknad på korleis språktrekket som vert undersøkt her, er realisert i den nye koinéen.

#### 1 Utgangspunktet for artikkelen

Denne artikkelen tek utgangspunkt i ein etablert, men også omstridd, hypotese innan koinéformingsteoriane. Dømet som blir nytta i denne artikkelen er Narvik, ein stad som hadde enorm folkevekst omkring år 1900. Ifølgje teoriane om koinéforming vil det då bli skapt ein ny dialekt, og mange teoretikarar meiner også at kva språktrekk denne nye dialekten får, i stor grad er avhengig av kvar innflyttarane kom ifrå. Dette kan formulerast som eit majoritetsprinsipp: Dei språktrekka som er i fleirtal blant den første bølga av innflyttarar til staden, vinn fram og dannar den nye dialekten (jf. Trudgill 2004). Intuitivt verkar dette logisk og riktig for mange, men som vi skal sjå er det omdiskutert, og det er heller ikkje lett å finne empiri som klart støttar ein slik hypotese – tvert imot viser

denne artikkelen at det ikkje er tydelege samanhengar mellom innflyttarane sine heimstaddialektar og utfallet av koinéformingsprosessen.

Artikkelen er bygd opp slik: Den byrjar med ei kort innleiing om dømet vårt, Narvik, og innflyttinga til denne staden (del 2). Dømet er valt fordi vi har så gode data om innflyttinga frå heilt riktig tidspunkt i historia, det vil seie folketeljningane frå 1900. I denne artikkelen utforskar eg majoritetsprinsippet ved å kombinere informasjon om fødestad i folketeljningane frå 1900 med talemålsdata frå målføresynopsisen. I del 3 forklarar eg kort koinéformingsteoriane og går særleg inn på majoritetsprinsippet. I denne delen trekker eg også vekslar på ein god del tidlegare forskning frå inn- og utland, for å vise korleis majoritetsprinsippet tidlegare har vore brukt som teori og forklaringsmodell. I del 4 presenterer eg målføresynopsisen, folketeljningane og korleis eg gjorde undersøkinga heilt konkret. Så kjem resultatata frå undersøkinga i del 5, medan del 6 gir ei drøfting av kva desse resultatata betyr og kva teoretiske implikasjonar eg meiner at dette bør få, og del 7 gir ei drøfting av målføresynopsisen som kjeldemateriale i historisk-sosiolingvistiske undersøkingar.

## 2 Dømet Narvik: Innflyttinga, skapinga av ein by og ein ny dialekt

Narvik er i dag ein stor kommune nord i Nordland og ein middels stor by i nordnorsk samanheng med om lag 14 000 innbyggjarar i byen og 22 000 i heile kommunen. Narvik oppsto som by rundt år 1900, då *Ofofbanen* vart bygd. Ofofbanen var eit gigantprosjekt initiert av gruveeigarar og myndigheiter som skulle utvinne jernmalm frå fjella i Kiruna i Nord-Sverige og transportere den med jernbane til Luleå i Bottenvika og Narvik i Ofoten. Der skulle malmen lastast på malmbåtar og transporterast vidare ut i verda. Narvik er såleis ikkje ein typisk industristad med ein stor fabrikk, men har hatt (og har) ei einsidig hjørnesteinsbedrift som driv med transport, lagring og lasting av jernmalm i industriell skala.

I 1887 byrja det første forsøket med bygging av Ofofbanen, men utbyggjarane gjekk konkurs i 1889 og dei 580 arbeidarane som framleis var til stades, vart arbeidslause (Aas 2001: 39). I 1898 byrja neste forsøk, og i 1902 vart Ofofbanen offisielt opna. Bygging av jernbane var på denne tida eit manuelt og sær arbeidskrevjande prosjekt: Om lag 5000 mann arbeidde på jernbaneanlegget på det meste (Frank 1950: 33). Narvik var ikkje eingong ei lita bygd før dette. Narvikhalvøya var buplass for nokre



Illustrasjon 1: Kart over Ofotbanen i Noreg og Sverige (Grytås 2017)

få gardbrukarar og låg på den andre sida av fjorden for det dåverande kommunesenteret Ankenes. I den første byggeperioden i 1880-åra hadde ein del arbeidrar (med familiar) kome tilflyttande, men dei reiste igjen då arbeidet vart lagt ned, så det budde ikkje mange der då det eitt tiår seinare vart klart at jernbanen skulle byggast ferdig og at endestasjonen for jernbanen og malmkaia skulle vere på Narvikhalvøya. Igjen fekk staden stor tilflytting, og denne gongen vaks Narvik fram som by. I 1898 budde det om lag 300 innbyggjarar på staden. På eitt år vaks innbyggartalet til om lag 2000 og i den offisielle folketeljinga frå 1900 er det 3705 innbyggjarar i byen (Det statistiske Centralbureau 1902: 49; Aas 2001: 50).<sup>1</sup> Det er ein ekstremt rask auke i folketalet på kort tid: meir enn ei tidobling på to år (sjå også figur 1).

Byen var ikkje ein planlagd by – verken frå den norske/svenske staten eller jernbaneutbyggerane si side (i motsetnad til mange andre norske industristader som vart planlagde og bygde av industrieigarar på denne tida

1. Narvik låg altså i utgangspunktet i Ankenes kommune og vart oppretta som bykommune i 1902. I 1974 vart Narvik og Ankenes slegne saman under namnet Narvik kommune. Når eg her skriv om Narvik, meiner eg busettinga på narvikhalvøya som vart bykommunen Narvik i 1902.

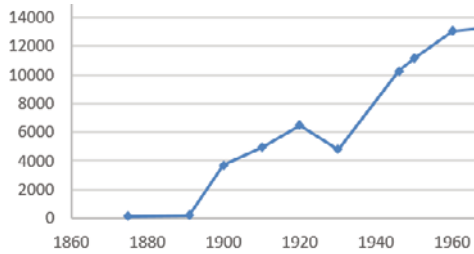


*Illustrasjon 2: Narvik i 1899. Bilete frå Museum Nord sine samlingar.*

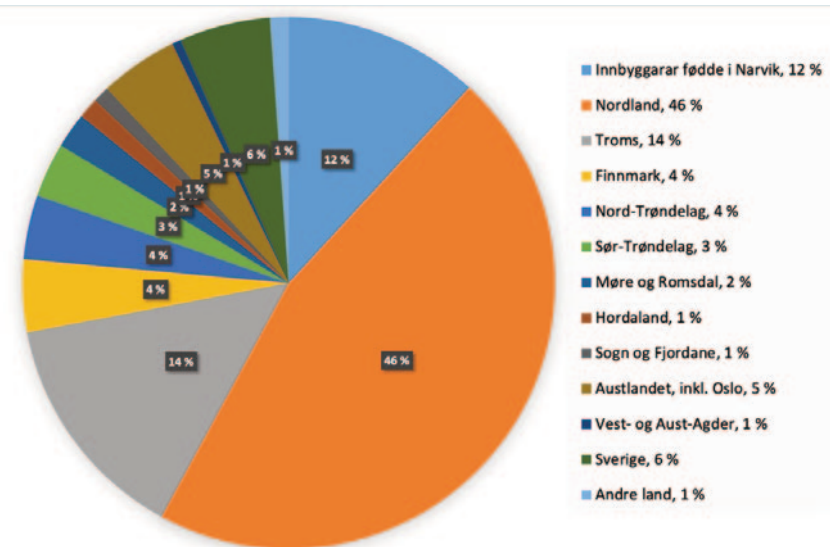
(jf. Utter 2010: 63–69)), og med så stor folkeauke på kort tid, vart det både bustadmangel og mangel på grunnleggande offentlege tilbod som reint vatn, vegar, renovasjon, helsehjelp og skule. Likevel viser folketelj- inga frå 1900 at mange av jernbanearbeidarane tok med seg familiane sine. Sjølv om det er ei overvekt av menn i befolkninga på dette tids- punktet (to tredelar er menn, jf. figur 3), finn vi også mange kvinner og barn, og heile 137 barn vart faktisk fødde i byen dette året. Det var (og er) tradisjon i Nord-Noreg for at mennene reiste vekk frå heimstaden på arbeid, til dømes på sesongfiske, medan kvinnene og borna var heime. Ut frå folketeljinga frå Narvik ser det ut til at mange av dei som kom til Narvik for å jobbe, ikkje såg på jernbanearbeidet som ein slik type mid- lertidig og tidsavgrensa jobb, men at dei tok med seg familien og busette seg i den nye byen. Folketalet fortsette å auke og i 1910 var det 4922 inn- byggarar i byen (Det statistiske Centralbyraa 1912).

Figur 2 gir ei grov oversikt over kvar innflyttarane til Narvik kom frå. Denne oversikta baserer eg på folketeljinga frå 1900 frå Ankenes kommune, og desse tala inkluderer både dei som budde i den nye byen på Narvikhalvøya, dei som budde i brakker langs jernbanen, og dei som budde i den midlertidige busettinga Rombaksbotn. Eg har valt å inkludere desse innbyggarane som budde utanfor bykjernen, men som var in-

## MED HEIMSTADDIALEKTEN I BAGASJEN?



Figur 1: Folketalsutviklinga i Narvik, basert på tal frå statistisk sentralbyrå (2021).



Figur 2: Innbyggjarane i Narvik sine fødestader. (Basert på fødestad oppgitt i folketeljinga frå Narvik i 1900.)

volvart i oppbygginga av jernbanen og byen, fordi desse må sjåast på som relevante for danninga av den nye dialekten (sjå meir om dette i del 4).

Innflyttarane til Narvik kom i all hovudsak frå Nordland (46 %) og Troms (14 %). I tillegg til innflyttarar frå Noreg kom det også ein del svenskar (6 %) og folk frå andre land (1 %). Legg merke til at det var fleire svenskar enn finnmarkingar i Narvik i 1900. Men det er kommunenivået som ligg til grunn for undersøkinga som blir presentert i denne artikkelen, og då er det snakk om kommuneinndelinga slik den var før 1950 (sjå



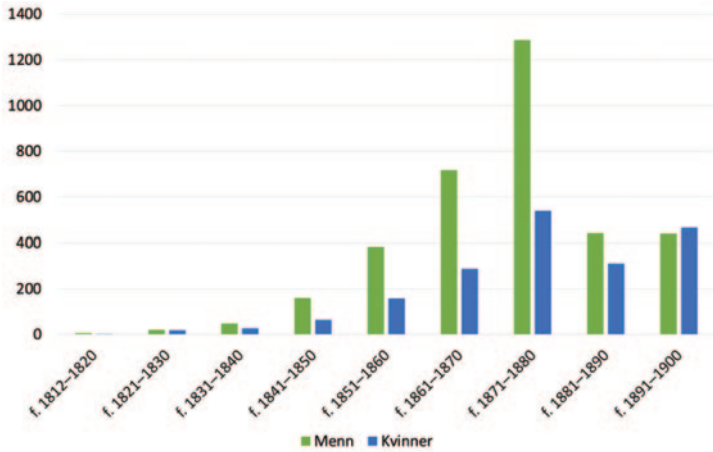
*Illustrasjon 3: Kart over kommunar i Nord-Noreg med særleg mange innflyttarar til Narvik. (Basert på fødestad oppgitt i folketeljinga frå Narvik i 1900.)*

meir om det metodiske i del 4). Frå Nordland kom særleg mange frå Lofoten (Vågan, Buksnes og Borge, no Vågan og Vestvågøy kommune), Evenes (i Ofoten), Mosjøen (kommunane Vefsn og Mosjøen, no Vefsn kommune) og Bodø (kommunane Bodø, Bodin og Skjerstad, no Bodø kommune). Frå Troms kom særleg mange frå Harstad (Trondenes kommune) og Målselv (sjå kart i illustrasjon 3).

Som nemnt ovanfor var det ei overvekt av menn blant innbyggerane i 1900 (to tredjedelar). Dette skil seg tydeleg frå den norske befolkninga elles, der det var 52 % kvinner (og 54 % kvinner i byane) (Det statistiske Centralbureau 1903: 2–3). Som vist i figur 3 var det også stort sett unge folk som flytta til den nye byen. I folketeljinga i 1900 var heile 64 % av innbyggerane i Narvik under 30 år. Dette er ei yngre befolkning enn landet som heilskap, sjølv om folketeljingane vitnar om stor folkevekst dei siste tiåra. 54 % av befolkninga var yngre enn 25 år (53 % i byane) og 35 % var under 15 år (32 % i byane) i Noreg i 1900 (Det statistiske Centralbureau 1903: 2–3).

Ut frå denne gjennomgangen ser ein at folketalet auka enormt raskt, mest med folk frå landsdelen, og at dei som flytta inn stort sett var unge folk. Ein slik enorm folkeauke på kort tid gjer det nesten umogleg å sjå for seg at dialekten kan vidareførast frå generasjon til generasjon på «van-





Figur 3: Aldersgrupper og kjønnsfordeling i Narvik. Basert på folketeljinga i 1900.

leg» vis. Denne situasjonen er typisk for det ein gjerne kallar koinéforming, der ein intens kontaktsituasjon mellom ulike gjensidig forståelege talemål, fører til at ein ny dialekt, ein koiné, blir danna (Kerswill 2010; Kerswill og Trudgill 2005; Neteland 2013). I denne artikkelen tek eg narvikdialekten som utgangspunkt for ein prinsipiell diskusjon om kva det er som avgjer korleis den nye dialekten blir. Sjå Neteland (2019) for eit oversyn over endringane i narvikdialekten og korleis den lokale koinéformingsprosessen gjekk føre seg.

### 3 Kva avgjer korleis den nye dialekten blir? Teori og tidlegare forskning

Koinéforminga blir oftast framstilt som ein tredelt prosess: Først kjem ein *kontaktfase*, der kontakten mellom talemåla oppstår og talarane akkommoderer mot kvarandre sine talemål. Deretter kjem ein *kaosfase*, der dei som vaks opp med særskilde mykje variasjon i innputten også snakkar med eit talemål som er kjenneteikna av mykje både interindividuell og intraindividuell variasjon. Etterkvart fører prosessen med seg språklege utjamningar og grammatiske forenklingar som gjer at den språklege va-

riasjonen blir mindre.<sup>2</sup> Då snakkar ein gjerne om ein tredje fase, *stabiliseringsfasen*. Til slutt i denne tredje fasen har koinéen blitt stabil og ein ny dialekt (ein koiné) er skapt (sjå t.d. Kerswill 2010; Kerswill og Trudgill 2005; Trudgill 1986; Neteland 2013).<sup>3</sup> Ein slik tredelt prosess er mellom anna dokumentert i dei norske industristaddialektane i Høyanger (Solheim 2006), Sauda og Årdal (Neteland 2014), i den engelske nye byen Milton Keynes (Kerswill og Williams 2000) og i engelsken som blir brukt på New Zealand (Trudgill 2004; Gordon m.fl. 2004). Desse tre fasane fell ofte saman med tre generasjonar, slik at det er den tredje generasjonen som veks opp på den nye staden som snakkar den nye dialekten (t.d. Solheim 2006; Neteland 2014).

Modellen av koinéformingsprosessen ser ut til å vere ei god skildring av *kva* som skjer når ein ny dialekt blir skapt. Men modellen er nettopp ei skildring og gir oss få forklaringar på *kvifor* koinéen blir akkurat slik den blir. Tenker vi for eksempel på dei som vaks opp i Narvik rundt år 1900, så veit vi at dei hadde få moglegheiter til å høyre den lokale dialekten – i staden for hørde dei eit mylder av ulike talemål rundt seg. Mange ulike språklege variantar var i omløp (jf. omgrepet ‘feature pool’ introdusert til sociolingvistikken av Mufwene (2001: 4–6)), og på eitt eller anna vis, vann ein av desse variantane fram og vart del av den nye dialekten. Men *kva* var det som avgjorde *kva* for ein av dei språklege variantane som vann fram, og *kva* variantar som tapte?

Heilt grovt kan vi sortere svara på desse spørsmåla i to; dei som vektlegg sosiale faktorar og dei som vektlegg lingvistiske faktorar. Dei sosiale faktorane kan til dømes vere statusen til varietetane som kjem i kontakt (t.d. Solheim 2006, som vektlegg standardtalemålet sin status som ei forklaring på utfallet av koinéforminga i Høyanger) eller kontaktmønster i det lokale språksamfunnet (som t.d. Schreier 2014, som legg vekt på barna sin sentrale rolle i koinéforminga når han forklarar kvifor mødrener talemål har fått størst innverknad på resultatet av koinéforminga av sør-atlantisk engelsk). Dei lingvistiske faktorane kan handle om språk-

2. Det finst mange definisjonar på «grammatisk forenkling». Eg baserer meg her på Kerswill og Trudgill si forklaring av omgrepet «simplification» som ein reduksjon av irregularitet i talemålsgrammatikken, for eksempel tap av kategoriar (som genus), færre bøyingsklassar eller eit mindre foneminventar (2005: 198).
3. Ein kan sjølvstøtt diskutere kor «stabil» ein dialekt er, sidan også ein «vanleg» dialekt vil ha ein heil del variasjon, men poenget her er at den ekstreme variasjonen som finst i byrjinga av koinéforminga, er redusert til «vanleg» dialektvariasjon.

struktur (t.d. at det i mange tilfelle er den grammatisk enklaste varianten som vinn fram, jf. Neteland 2019: 52–54) eller om markertheit og frekvens (t.d. at dei variantane som er mest frekvente, har stor sjanse for å vinne fram, jf. Trudgill 1986: 102). Det er denne siste forklaringsmodellen som er utgangspunkt for drøftinga i denne artikkelen.

Frekvens har i mange tilfelle blitt nært knytt til demografien på den nye staden, altså til samansetjinga av innflyttarar og kva talemål dei hadde med seg i bagasjen. Ein tenker seg at når mange dialektar møtest, er det talemålet til dei innflyttarane som er i majoritet, som har størst sjanse for å vinne fram. Dette har vore ei ganske vanleg og lite kontroversiell forklaring både i norsk og internasjonal forskingslitteratur. For eksempel blir det hevda i publikasjonar frå 1950-talet til i dag, at dialektane i naboindustriestadene Odda og Tyssedal vart så forskjellige fordi dei vart befolka av innflyttarar med forskjellige dialektar (Sørli 1959: 56; Sandve 1976: 171; Kerswill 1994a: 224; 2013: 523–525; Sandøy 1985: 250–251; 2000: 357–359). Som eit døme på kor vanleg og ukontroversiell denne forklaringa er, vil eg trekke fram eit sitat frå den nye norske språkhistoria:

“Innflyttingskontingentane” frå ulike dialektområde i Noreg var altså nokså ulike i Odda og Tyssedal. Derfor blei det språklege resultatet at den nye tyssedalsdialekten var nokså prega av søraustlandsk, noko som ikkje er heilt uventa når 33,8 % kom frå Sør-Austlandet og 28 % frå andre ikkje-vestlandske område. Oddamålet derimot heldt fram med å vere hardingmål, berre noko forenkla i fonologi og morfologi, for her kom over 60 % av tilflyttarane frå Vestlandet. (Sandøy og Nesse 2016: 67)

Ifølge denne forklaringa er det altså proporsjonane av innflyttarar med ulike dialektar som legg grunnlaget for korleis den nye dialekten blir. Forklaringa har ikkje berre blitt brukt om industristadtalemåla Odda og Tyssedal, men også for å forklare kvifor årdalsmålet vart såpass lite endra etter industrialiseringa (Bjørkum 1974: 548) og er framstilt som ein av fleire samverkande faktorar for danninga av det nye talemålet i Høyanger (Solheim 2006: 388–391; Omdal 1979: 79; Trudgill 1986: 98). Også forskarar som jobbar med andre språkområde enn det norske, peikar på proporsjonar av innflyttartalemål som ein vesentleg faktor i koinéformingsprosessane (Mufwene 2001: 4–6; Trudgill 1986: 98–107; Siegel 2001: 179; Britain 2012: 223–231). Men då Trudgill i 2004 brukte resultatata frå forskingsprosjektet sitt om engelsk på New Zealand til å

sette fram påstanden om at innflyttarane sine talemål *determinerer* utfallet av koinéforminga, vart det debatt (sjå særleg Trudgill 2008a og svara i *Language in Society* 37(2) 2008, og også Kerswill 2007; 2010).

Nokre av argumenta mot denne påstanden er at frekvens berre er éin av fleire faktorar som avgjer kva variant som vinn fram (Mufwene 2008: 19–20) eller at nokre innflyttarar er viktigare enn andre for utfallet av prosessen (Schreier 2014). Nokre av tilsvara til Trudgill la også vekt på at andre sosiale faktorar spelar inn på koinéformingsprosessen (Kerswill 2010). Særleg vart identitet trekt fram (Mufwene 2008; Tuten 2008; Schneider 2008; Coupland 2008; Bauer 2008; Holmes og Kerswill 2008), busetnads- og kontaktmønster mellom innbyggerane (Hickey 2003: 221; Mufwene 2006: 179) eller at språkleg ideologi kan gjere at prestisjetalemål eller -talarar har større sjanse for å vinne fram (Solheim 2009: 199; også diskutert i Solheim 2006: 85–89). Trudgill svara på denne kritikken med å seie at demografien berre kunne føreseie utfallet av koinéformingsprosessen i såkalla *tabula rasa*-situasjonar, altså viss eit heilt nytt samfunn blir forma og dei nye innbyggerane (altså barna til emigrantane) er isolert frå det samfunnet dei reiste frå, slik at dei «gamle» sosiale normene, prestisjen osv. ikkje lenger gjer seg gjeldande (Trudgill 2008b: 279; også diskutert i Trudgill 2004: 26). Dette argumenterer Trudgill for at var tilfellet i New Zealand, og den språklege empirien han presenterer, er overtydande: Proporsjonane av innflyttarar frå ulike stader av Storbritannia blir gjort om til proporsjonar av språklege variantar som innflyttarane hadde med seg, og denne mengda av ulike variantar blant innflyttarane speglar elegant kva variantar som vann fram og kva variantar som måtte gje tapt i løpet av koinéformingsprosessen (Trudgill 2004: 113–128).

Ein annan grunn til at Trudgill sin hypotese er god, er at den er mogleg å teste, i motsetnad til dei fleste sosiale faktorene som ein berre kan sannsynleggjere og argumentere for eller mot på bakgrunn av teori og lokalkunnskap. I Trudgills majoritetsprinsipp kan ein, viss ein har tilstrekkeleg kunnskap om kvar innflyttarane kom frå og om dialektane som vart snakka der, rekne ut kva språkleg variant som var i majoritet i innflyttingsperioden, og dermed også føreseie kva variant som til slutt ville vinne fram i koinéformingsprosessen. I denne artikkelen har eg valt å gjennomføre ei slik utrekning, til tross for at danninga av Narvik by var langt frå ein *tabula rasa*-situasjon. Dette er fordi demografi og innflyttarane sine talemål har vore ein vanleg forklaringsmodell i norske so-

siolingvistiske granskingar av industristadtalemål (sjå ovanfor), og eg ville difor undersøke i kor stor grad innflyttarane sine talemål kunne forklare utfallet av koinéformingsprosessen. For å kunne gjere ei slik utrekning av kva språktrekk som var i majoritet blant innflyttarane, må ein ha nøyaktige data om kvar innflyttarane kom frå, og når ein jobbar med norske talemål, er det ikkje nøyaktig nok å ha innflytterdata på fylkesnivå, fordi norske talemål har så stor lokal variasjon. Dette har eg prøvd før (Neteland 2014), og eg trur den artikkelen viser godt kor problematisk det er å bruke innflytterdata frå fylkesnivå – for eksempel er det umogleg å seie kva slags bøyingsformer av substantiv i hokjønn innflyttarar frå Hordaland hadde med seg, både fordi bergensk har felleskjønnsystem og fordi dei andre dialektane i fylket har mange ulike bøyingsendingar (jf. Neteland 2014: 34–36). I denne artikkelen har eg nøyaktige data om innflyttarane sine fødestader på kommunenivå frå folketeljinga, og eg brukar målføresynopsisen som empirisk dokumentasjon av dialektane på stadane dei kom frå. På denne måten kan eg gjere presise utrekningar av kva språktrekk som var i majoritet blant innflyttarane, og slik gjennomfører eg ein test av Trudgills hypotese om at innflyttarane sine fødestadstalemål determinerer utfallet av koinéformingsprosessen.

#### 4 Kjeldemateriale og framgangsmåte

Til denne undersøkinga er det to hovudkjelder: *Folketeljinga frå 1900*, som er brukt til å finne fødestaden til alle innbyggerane i Narvik på det tidspunktet, og *målføresynopsisen*, som er brukt til å finne ut kva språkleg variant innbyggerane hadde i fødestadstalemålet sitt. Ut frå denne kombinasjonen har eg rekna ut kva språkleg variant som var i fleirtal blant innbyggerane som budde i Narvik i år 1900.

##### 4.1 Målføresynopsisen

Ein 'synopsis' er ei parallelloppstilling av eit emne, og målføresynopsisen er ei oversikt over dialekttrekk frå heile Noreg. Synopsisen er basert på spørjeordlista som er best kjend som *Storms ordliste* (Storm 1884), og gir informasjon om fonologi og morfologi og nokre få syntaktiske konstruksjonar. Dei empiriske dataa til synopsisen vart samla inn i perioden 1930–1970, mest av alt i perioden 1945–1960 (Hoff 1960: 596–598). Sy-

nopsisen er geografisk sortert etter dei gamle fylka og kommunane<sup>4</sup> og samla i 43 bind.<sup>5</sup> Kvar region har sitt band, og i bandet står kvart talemål omtalt på kvar si rad (horisontalt) og kvart språktrekk har kvar si kolonne (vertikalt). Synopsisen kan såleis brukast som forskingsressurs anten ein les horisontalt eller vertikalt. Horisontal lesing gir informasjon om fonologien og morfologien i kvar enkeltdialekt og dette gir eit godt utgangspunkt for vidare undersøkingar av variasjon og endring på staden. Vertikal lesing gir moglegheit til å undersøke distribusjonen av eitt språktrekk i ein region eller i hele landet.

Synopsisen er laga på eit seinare tidspunkt enn det eg er oppteken av i denne undersøkinga. Men sidan synopsisen vart samla inn etter tradisjonelle dialektologiske metodar, der eldre informantar med lokalforankra familie og lite språkkontakt utanfor lokalmiljøet var føretrekt (Hoff 1960: 602; Wetås 2014: 319), er det gode grunnar til å tru at talemålsdataa den framstiller, ikkje representerer eit talemål som var moderne på innsamlingstidspunktet, men heller eit talemål som var tradisjonelt allereie då. Tenker vi oss til kven som var «eldre» omkring 1950–1970-talet, kjem vi nettopp til dei som var unge og unge vaksne i år 1900. Med støtte i hypotesen om tilsynelatande tid, der ein reknar med at eldre informantar representerer dialekten slik den var då dei var unge vaksne (Labov 1994: kap. 3, særll. 45–46), vil eg difor argumentere for at desse talemålsdataa er ein god representasjon av talemåla som vart brukt på heimstadene som innflyttarane til Narvik hadde flytta frå. Målføresynopsisen er i det heile teke eit verdifullt datamateriale for sosiolingvistiske undersøkingar med ei historisk vinkling, så lenge ein tek høgde for at materialet ikkje er samla inn med moderne sosiolingvistiske metodar og idear om representativitet.

4. Altså kommunenamna slik dei var før dei omfattande kommunesamanslåingane i kjølvatnet av innstillinga til Schei-komiteen i 1952 og fylkesnamna slik dei var før samanslåinga i 2020. I ein del av kommunane er dessutan talemålet registrert på eit par forskjellige geografiske nedslagsfelt.
5. Slik eg forstår det, var det tenkt at synopsisen skulle vere grunnlag for eit nasjonalt dialektatlas som til dømes *The Linguistic Atlas of England* (Orton, Sanderson og Widdowson 1978) og *The Atlas of North American English* (Labov, Ash og Boberg 2006). Eg veit ikkje kvifor dette ikkje vart publisert, men det finst ein god del dialektkart som er laga på bakgrunn av målføresynopsisen. Bakgrunnsdataa, desse karta og sjølve målføresynopsisen finst i Språksamlingane sitt arkiv og er også digitaliserte. Språksamlingane held også på med å lage ein ny og moderne versjon for søk i målføresynopsisen som skal vere klar før denne artikkelen blir publisert. Sjå <https://synopsis.uib.no>.

#### 4.2 Val av språkleg variant

Majoritetsprinsippet går altså ut på at den språklege varianten som er i fleirtal blant innbyggerane i etableringsfasen, er den som har størst sjanse for å vinne fram i koinéformingsprosessen. Det er mange språktrekk ein intuitivt kan sjå at gir støtte eller ikkje støttar majoritetsprinsippet når ein legg til grunn at folk snakka fødestadtalemålet sitt på den nye plassen. For Narvik sin del er det slik at sidan dei fleste innflyttarane kom frå Nord-Noreg (jf. figur 2 og 3), er det openbert at den nye dialekten burde ha hatt palatalisering av alveolarar og at *ikkje* skulle ha vore brukt som nektingsadverb, viss majoritetsprinsippet hadde vore styrande for utviklinga, men desse variantane er ikkje brukt i den stabiliserte koinéen. Samstundes kan det sjå like openbert ut at forklaringa på at koinéen i Narvik brukar dei vanlegaste nordnorske fleirtalsbøyingane av hokjønns substantiv (*-e* og *-en*) og hankjønns substantiv (*-a* og *-an*), er at dei fleste innflyttarane kom frå Nord-Noreg. Desse variantane vann altså fram i koinéformingsprosessen, medan variantar som *ikkje* og palatalisering ikkje gjorde det, sjølv om fleirtalet av innflyttarane må ha hatt med seg alle desse språktrekka i bagasjen (Neteland 2019: 23–25).

Desse døma viser tydeleg at det er noko som «skurrar» med majoritetsprinsippet, og i denne artikkelen ønskte eg å setje prinsippet på prøve med nøyaktige utrekningar av kva variant som var i majoritet blant innbyggerane sine fødestadstalemål. For min eigen del vart dette meir interessant dersom det ikkje var så openbert kva variant som var i majoritet. Eg valde difor ending på svake hokjønnsord i ubestemd form eintal, som varierer mykje i Nord-Noreg. Språktrekket varierer også i koinéen i Narvik mellom *-e* og *-a* (til dømes *ei kanne/ei kanna*), der *-e* er den dominerande varianten som blir brukt i om lag 80 % av tilfella, medan *-a* blir brukt i om lag 20 % av tilfella (Neteland 2009: 76). Desse tala gjeld både for ungdoms- og eldregruppa i materialet frå 2008 (født i 1990 og på 1920-talet), så variasjonsmønsteret er difor sannsynlegvis ikkje eit uttrykk for endring, men dataa viser ein heil del individuell variasjon og variasjon langs sosiale skiljeliner (Neteland 2009: 76–77). Det er difor interessant å sjå om dette variasjonsmønsteret kan ha samanheng med innflyttarane sine talemål.

I målføresynopsisen (og Storms ordliste) er svake hokjønnsord i ubestemt form eintal representert i paragraf nr. 37 med orda *ei vise, ei jente, ei side, ei kiste, ei kyrkje, ei kanne, ei panne, ei oske, ei vogge, ei veke, ei fure, ei stove, ei gate, ei hose, ei kone, ei reke, og ei stor glede* (jf. Storm 1884: 6–7).

Synopsisen har ikkje data som seier noko direkte om talemålet i Narvik sentrum, men frå dataa frå bygdene omkring Narvik er det klart at det er *-a* (til dømes *ei kanna*) som har vore brukt i den tradisjonelle dialekten i området.<sup>6</sup> Variasjonen er elles ganske stor i Nord-Noreg: I tillegg til at *a*-endinga er brukt fleire stader (t.d. Hammerfest og Vågan), finn vi også *e*-ending mange stader (t.d. Målselv og Alta) og ein djupare artikulert *æ*-variant som også er notert *ä* i synopsisen (t.d. Rana og Vefsn). Synopsisen viser også at det er variasjon mellom *e*- og *a*- nokre stader (t.d. Lødingen). I dei områda der det er apokope, viser målføresynopsisen at svake hokjønnsord i ubestemt form eintal får apokopert ending (t.d. Bodø og Buksnes). Desse variantane er også utbreidde blant dei trøndske innflyttarane til Narvik. I tillegg kjem jamvektsdialektane, som er mest utbreidde blant innflyttarar frå Trøndelag og Austlandet. Desse har anten apokope og jamvekt (t.d. Bindal og Namsos) eller *-e* og jamvekt (t.d. Røros) i svake hokjønnsord i ubestemt form eintal.

#### 4.3 Framgangsmåte ved registrering av fødestad og av språkleg variant

Heilt konkret leita eg opp folketeljinga frå Ankenes kommune frå 1900 (Digitalarkivet 2000 [1900]) og la den inn i eit excelark med ei rad for kvar person. Deretter fjerna eg dei personane som budde i dei rurale områda av kommunen, og såleis ikkje var like sentrale i skapinga av den nye dialekten som innbyggerane i byen og rallarane som jobba med bygginga av jernbanen. Eg stod då att med 5482 personar som budde i byen, i brakker langs jernbaneanlegget og i den midlertidige busettinga i Rombaksbotn.<sup>7</sup>

Deretter leita eg opp fødestaden til desse personane i målføresynopsisen og fann ut kva variant som var brukt der. Det var ikkje alle innbyggerane som hadde oppgitt fødestad i folketeljinga (11 personar mangla fødestad), og for nokre var det berre oppgitt fylke (35 personar). Eg fann heller ikkje informasjon om variant frå alle stadene i målføresynopsisen (47 personar). Dessutan var ein del av innflyttarane frå Sverige (302 personar) og andre land (62 personar). Likevel fann eg informasjon om dette språktrekket for totalt 5038 personar ved hjelp av målføresynopsisen.

6. I synopsisen er det talemålet i Skjomen som er oppgitt for kommunen. Der er det *a*-ending på svake hokjønnsord i ubestemt form eintal. Det er også *-a* i nabokommunane Ballangen og Evenes.
7. Rombaksbotn vart fråflytta då jernbanen var ferdig bygd i 1902, og både folk og bygningar vart flytta til Narvik.



Dei ulike variantane vart til slutt rekna saman, slik at eg fann proporsjonane av varianter blant innbyggerane sine fødestadstalemål på dette tidspunktet.

#### 4.4 Nokre val

Sjølv om dei fleste innbyggerane var innflyttarar, var det også ein del som var fødde i kommunen (649 personar). Desse forvanskar undersøkinga, sidan koinéforminga i Narvik då ikkje er ein «tabula rasa»-situasjon: Det er både enkel og utstrekt kontakt med det større språksamfunnet og kontakt med tradisjonelle dialektbrukarar på lokalt nivå. Eg vurderte det likevel slik at desse skulle inkluderas i materialet, sidan det er slik det har vore gjort i tidlegare undersøkingar av koinéformingsprosessar i Noreg, t.d. i Odda og Årdal (jf. Sandve 1976; Bjørkum 1968).

Ei anna problemstilling som dukka opp, var at så mange som 166 innbyggerar var fødde i år 1900 og altså var under eitt år på dette tidspunktet. Ein eittåring kan som regel berre seie nokre få ord (Simonsen m.fl. 2013: 12), og ein kan dermed sjå det slik at desse spedbarna ikkje bidreg med noko særleg i dette første stadiet av koinéforminga. Eg valde likevel å inkludere dei i undersøkinga med fødestaden som var oppgitt for dei, sidan det både er høgst usikkert og vanskeleg å operasjonalisere ei aldersgrense for kor gammalt eit barn skal vere før det bidreg til skapinga av ein ny dialekt. Dei fleste av desse spedbarna var fødde i Ankenes kommune (137 eittåringar), og det er såleis særleg varianten som målføresynopsisen oppgir for kommunen (-a) som blir noko større når desse spedbarna er inkludert i materialet.

Det er også, som nemnt ovanfor, 364 utlendingar blant innflyttarane. Særleg svenskane utgjer ei såpass stor gruppe (302 personar, dvs. om lag 6 % av folketalet) at det gjer store utslag om ein tek dei inn i utrekningane eller ikkje. Vi har ikkje empirisk dokumentasjon på korleis innvandrarar til Noreg snakka rundt år 1900, men vi veit at svenskane kunne gjere seg forstått med å snakke svensk. Sidan mange av svenskane ikkje har oppgitt fødestad (berre Sverige) og ein del har oppgitt *län* (altså fylke), er det heller ikkje mogleg å gjere ei like nøyaktig inndeling av fødestadstalemålet deira som av nordmennene i undersøkinga. Sidan undersøkinga elles er basert på empirien frå målføresynopsisen, blir det ikkje ei god løysing å estimere kva slags svensk eller norsk innflyttarane frå andre land snakka. Eg valte difor å ekskludere svenskane og dei andre utlendingane frå utrekningane.

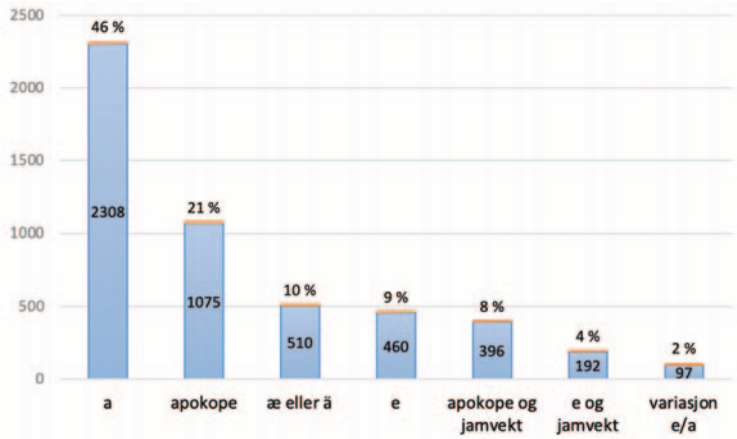
## 5 Resultat av undersøkinga

Kva ending svake hokjønnsord har i ubestemt form eintal, varierer som nemnt mykje mellom dei nordnorske dialektane, og i målføresynopsisen ser ein tydeleg korleis dette varierer på lokalt nivå frå kommune til kommune. Dette har eg altså kombinert med informasjon om innflyttarane sine fødestader frå folketeljinga i 1900. Figur 4 summerer opp resultatata av undersøkinga grafisk med ein stolpe for kvar variant, heiltal inne i stolpen og prosentdel på toppen. Blant innbyggjarane i Narvik i 1900 sine fødestadstalemål er det endinga *-a* som klart dominerer (*ei flaska*), deretter kjem dialektane med apokope (*ei flask*), og dialektane med *-æ* og *ä*-ending (*ei flaskæ*, *ei flaskä*), tett følgt av *-e* (*ei flaske*).

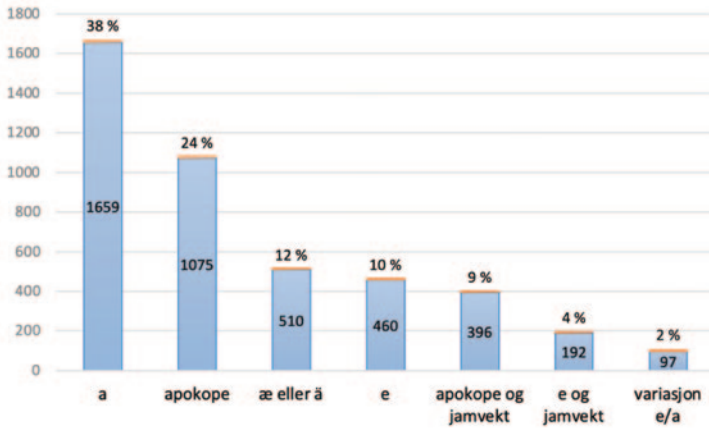
Som nemnt ovanfor er også lokale innbyggjarar frå Ankenes kommune inkludert i tala for *-a*, men sjølv om ein fjernar desse frå dataa, vil *-a* vere den klart mest frekvente varianten (sjå figur 5).

Sidan dette handlar om frekvens av variantar, kan ein også tenke seg at jamvektskategoriane bør slåast saman med den varianten som jamvektsformene er kombinert med, sidan desse er mest frekvente. Altså at ein tek dei dialektane som vekslar mellom apokope og jamvektsendingar, inn i kategorien *apokope* (t.d. Bindal), mens dei dialektane der endingane vekslar mellom *-e* og jamvektsending, (t.d. Røros) går inn i kategorien *-e*. Med slike grep er det særleg apokope-gruppa som veks, men heller ikkje desse utrekningane endrar på hovudtendensen i datamaterialet: *-a* er uansett den mest frekvente varianten, etterfølgt av apokope på andre-plass (sjå Neteland 2019: 68).

I koinéen i Narvik er det som tidlegare nemnt *-e* som dominerer (om lag 80 %), medan *-a* blir brukt i om lag 20 % av tilfella både blant unge og eldre informantar. Som ein kan sjå, føreseier altså *ikkje* talemåla til innflyttarane utfallet av koinéformingsprosessen for dette språktrekket. I så fall skulle det jo ha vore *-a* som dominerte, og opnar ein opp for variasjon, måtte ein kunne forvente at variasjonen var mellom *-a* og apokope, sidan også andelen innbyggjarar med apokope i fødestadstalemålet er ganske høg. Apokope i svake hokjønnsord i ubestemt form eintal finst det ingen døme på i materialet mitt frå Narvik (Neteland 2009; 2019). Endinga *-e* i ubestemd form eintal av svake hokjønnsord er heller ikkje noka grammatisk forenkling, så det er heller ikkje ei klår språkintern forklaring på at *-e* vann fram.



Figur 4: Tal på variantar i fødestadstalemåla til innbyggerane i Narvik i 1900.



Figur 5: Tal på variantar i fødestadstalemåla til innflyttarane til Narvik i 1900, utanom innbyggerar som er fødte i kommunen.

## 6 Drøfting av resultatata

Med hjelp av folketeljinga frå 1900 og målføresynopsisen kan vi vise at innflyttarane sine fødestadstalemål ikkje er ei direkte årsak til at koinéen i Narvik vart akkurat slik den vart. Eller sagt med andre ord: Fødestadstalemåla til innbyggerane determinerte ikkje utfallet av koinéformings-

prosessen. Det kan såleis vere freistande å trekke konklusjonen at majoritetsprinsippet som Trudgill formulerte, bør avvisast og at vi heller bør sjå nærare på dei andre, alternative forklaringane nemnt i del 3. Men i så fall trekker vi konklusjonen litt for langt, for det eg har undersøkt, er ikkje *verkeleg* dokumentasjon av korleis innflyttarane snakka – eg har laga ein rekonstruksjon av språkpotten ('feature pool') der eg har basert meg på premissen om at innflyttarane snakka talemålet frå heimplassen sin når dei kom til den nye staden. Dermed kan det like godt vere denne premissen, denne operasjonaliseringa av majoritetsprinsippet, som er feil, ikkje sjølve hypotesen om at variantane som er i fleirtal i byrjinga av koinéformingsprosessen, er dei som til slutt vinn fram.

Denne operasjonaliseringa, der ein reknar med at innflyttarane til ein ny stad snakka fødestadstalemålet sitt, er ikkje ny: Dette er grunnlaget for Trudgill sine utrekningar, og mange norske forskarar legg også vekt på demografi (og då særleg kvar innflyttarane kom frå) når dei forklarar kvifor norske industristadstalemål vart slik som dei vart. Også ein del forskarar som meiner at det er mange faktorar som spelar inn på koinéformingsprosessen, går ut frå at innflyttarane snakka fødestadstalemålet sitt. Omdal (1979: 75) seier det slik: «I en slik målblendingssituasjon som en da fikk i Høyanger, må en kunne gå ut fra at innflytterne stort sett beholdt sitt opphavlige talemål livet ut, men at det før eller seinere blei utvikla en ny Høyanger-dialekt hos barna.» Han meiner altså at innflyttarar til ein industristad der det ikkje er noko etablert talemål, vil halde på heimstaddialekten sin, og ein slik tankegang er grunnlaget for mange av desse demografiske forklaringsmodellane.

Nettopp difor vil eg dvele litt ved dette poenget: Vi har ikkje dokumentasjon på at innflyttarane til industristadene snakka fødestadstalemålet sitt når dei kom til den nye staden, og heller ikkje på at dei fortsette med det livet ut. Forsking på innflyttarar til industristader viser at mange har budd mange andre stader før dei kom fram til industristaden – dei kom altså ikkje direkte frå heimstaden sin (Hagen 1953; Aarhun 1973).<sup>8</sup>

8. Det er sjølvstakt ikkje så lett å talfeste kor mange som har budd andre stader når ikkje folketeljingane gir informasjon om dette. Hagen (1953) såg på lister over dei tilsette ved smelteverket i Sauda, der arbeidarane hadde oppgitt både fødestad og tidlegare arbeidsstad. Her var det berre 209 av 716 som hadde kome direkte til Sauda (Hagen 1953: 24). Aarhun (1973) undersøkte flyttemønstra til innflyttarane til Sauda ved å sjå på familiar i folketeljingane. Ut frå kor ungane var fødde, kunne kan sjå at dei aller fleste familiane hadde familiemedlemmer som var fødde på ulike stader (berre

Forskning på dialektbruken til nordmenn som flyttar, viser at dei endrar på fødestadsdialekten sin (jf. t.d. Mæhlum 1986; Omdal 1994; Kerswill 1994b; Kleivenes 2002; Hernes 2006). Eg har sjølv undersøkt talemålet til innflyttarar til industristadene Sauda og Årdal (Neteland 2017; 2019: 78–80) og det mest slåande i dette materialet er at alle kan plasserast på ein skala mellom fødestadstalemålet sitt og eit standardtalemål. Det vil seie at dei ikkje brukar lokale variantar, verken nye eller tradisjonelle, med mindre dei same variantane også fanst i fødestadstalemålet deira eller i ein standardvarietet.

I dag finst ein forholdsvis sterk sosial norm i det norske språksamfunnet om å «halde på dialekten», som vi kan skildre som at ein helst ikkje skal endre på fødestadstalemålet sitt gjennom livet sjølv om ein flyttar til ein annan stad og etablerer seg der. Eg meiner at denne norma om å «halde på dialekten» er forholdsvis ny i det norske språksamfunnet, og at den konkurrerer med tidlegare sosiale normer om å tilpasse seg andre med å «normalisere» dialekten sin og å snakke standardtalemål. Dette seier eg både med grunnlag i forskinga på flyttarar, og med grunnlag i forskinga på den historiske bruken av talt riksmål i Noreg (Nesse 2015; 2008; Jahr 2013; 2008 [1984]; 2007; Sandøy 2011; Hoel 2018: 436–438; Rambø 2018: 520). Ifølge Jahr (2007: 94) nådde bruken av talt riksmål sitt høgdepunkt i Noreg i mellomkrigstida. Det var med andre ord aukande bruk av talt riksmål i samfunnet på den tida då Narvik vart etablert rundt år 1900. Talt riksmål vart, i likskap med standardtalemål i andre land, forventa brukt av visse personar (altså høgare sosiale lag av befolkninga) og i visse situasjonar, til dømes i offentlege kommunikasjonssituasjonar (kyrkje, teater, radio, og anna offentleg tale) (Nesse 2015), men også i høfleg tiltale til personar ein ville vise respekt, og elles i transaksjonar mellom framande (butikk, offentleg kontor osv.) (Nesse 2008: 51). Talt riksmål var difor ikkje berre ein sosiolekt, men også eit talemål som så å seie alle var forventa å kunne bruke i visse situasjonar.

Viss innflyttarane til Narvik i år 1900 ikkje var særskilt opptekne av å halde på dialekten sin og ikkje berre snakka fødestadstalemålet sitt, men også i mange tilfelle normaliserte dialekten sin og snakka riksmål, blir språkpoten heilt annleis samansett fordi bruk av riksmålsvariantar også må reknast med. Kor mykje og kor ofte kvar enkelt person brukte

i 10 av 82 familiar var alle fødde på same stad). Dei fleste hadde budd i andre byar og industristader før dei kom til Sauda (Aarthun 1973: 74–76).

riksmålsvariantar og variantar frå fødestadstalemålet sitt, ville vere avhengig både av sosial klasse, arbeidssituasjon, sosiale nettverk og personlegdom (Nesse 2015: 108), så nøyaktige utrekningar av korleis språkpoten i Narvik i 1900 faktisk var samansett, vil vere umogleg. Li-kevel trur eg dette er noko av svaret på problemet med majoritetsprinsippet. Det er ikkje nødvendigvis majoritetsprinsippet det er noko i vegen med, men operasjonaliseringa der ein set likskapsteikn mellom innflyttarane sin fødestad og talemålsbruken deira. For å kunne gjere det, treng vi inngåande sosio-historisk kunnskap om språkbruk i samfunnet på det aktuelle tidspunktet, og ideelt sett både kunnskap om språkbruken på lokalt nivå og på eit meir overordna, nasjonalt nivå. Trudgill argumenterer for at innflyttarane til New Zealand hadde så lite kontakt med andre språksamfunn at dei kunne byrje med blanke ark utan omsyn til språknormene som var operative der dei reiste frå ('tabula rasa'-situasjon). Eg stiller meg bak dei som spør om ikkje folk tok med seg språkideologiane sine (f.eks. om bruk av standardtalemål og høflegheit) på flyttelasstet og overførte dette til ungane sine (Holmes og Kerswill 2008; Kerswill 2010: 238). I ein norsk samanheng blir ein slik tankegang enno meir problematisk, sidan industristadene som vaks fram på byrjinga av 1900-talet, ikkje oppstod i noko sosialt vakuum, men var raskt veksande, moderne byar som var del av eit større samfunn. I dette større, nasjonale språksamfunnet var det etablerte språklege normer som eg meiner må ha påverka utfallet av koinéformingsprosessen – dei overordna normene skapte mønster for språkbruk også på dei nye stadene. Koinéforminga i Sauda, Årdal, Narvik, Odda, Høyanger og Tyssedal var dermed ikkje berre lokale prosessar der innflyttarane sine talemål blanda seg saman og etterkvart skapte ein ny dialekt, men prosessar som var påverka av den tida dei oppsto i, dei spesifikke lokalhistoriske forholda, og det større språksamfunnet som dei nye stadene var ein del av.

## 7 Drøfting av målføresynopsisen som kjeldemateriale

Resultatet av denne undersøkinga viser oss først og fremst at det ikkje er nok å sjå på fødestadene til innflyttarane til industristadene når vi vil forstå kvifor talemåla utvikla seg på den måten dei gjorde. Koinéformingsprosessen på industristader er langt meir kompleks enn det. Dette kunne vi ikkje ha dokumentert like tydeleg og presist utan datamaterialet

i målføresynopsisen. Målføresynopsisen er på ein gong både ei kjelde til informasjon om heile grammatiske system i dei enkelte talemåla våre, og ei kjelde til informasjon om distribusjonen av dei enkelte språktrekka som varierer i norsk. Med denne artikkelen har eg også vist at den detaljerte talemålsinformasjonen i målføresynopsisen kan kombinerast med folketeljningane eller andre historiske data viss ein vil talfeste frekvensar av ulike språklege variantar.

Utfordringane med å bruke målføresynopsisen er for så vidt dei same som fordelane: Den dokumenterer dei tradisjonelle dialektane, ikkje korleis folk i ulike sosiale lag snakka, og viser sjeldan fram om språktrekk varierte i dei enkelte dialektane. Dette gjer at datamaterialet i synopsisen er konsistent og samanliknbart frå stad til stad, men det inneber også at ein må ta høgde for at informasjon om sosial variasjon manglar når ein skal bruke den som grunnlag for undersøkingar. Målføresynopsisen sin vekt på tradisjonelle dialektar og den konsistente metodologien som er brukt på tvers av alle stadene – inkludert mangel av framvising av sosial variasjon – er ein fordel i undersøkinga som er presentert i denne artikkelen. Det er dette som mogleggjer testing av majoritetsprinsippet. At majoritetsprinsippet bygger på ein feil premis om at folk som flyttar til ein stad som ikkje har ein etablert dialekt, tek med seg fødestadstalemålet sitt utan å endre på det, kan vi ikkje klandre målføresynopsisen for. Eg meiner at målføresynopsisen er eit datamateriale som gir mange moglegheiter for historisk-sosiolingvistiske undersøkingar av talemål – og så må ein sjølv sagt ta omsyn til at synopsisen viser fram éi side av talemålsbruken som fanst i landet på denne tida.

## 8 Avrunding

I denne vesle undersøkinga har eg vist fram ein måte ein kan bruke målføresynopsisen som kjeldemateriale i historisk sosiolingvistik. Artikkelen tok utgangspunkt i ei teoretisk problemstilling om samanhengane mellom innflyttarane sine talemål og utfallet av koinéformingsprosessane, og eit konkret språktrekk (hokjønnsord i ubestemd form eintal). Som vi har sett, var ikkje den nye dialekten som oppstod i Narvik særleg prega av talemåla frå heim plassane til innflyttarane når det gjeld dette språktrekket. Dette gjeld også ein del andre utbreidde nordnorske talemålsvariantar (som palatalisering av alveolarar, sjå del 4.2). Eg har argu-

mentert for at det er sosio-historiske årsaker til at dei mest utbreidde variantane i innflyttartalemåla ikkje vann fram, mellom anna dei sosiale språkbruksnormene som fantes i det norske språksamfunnet som heilskap på denne tida, som eg meiner at også verka inn på språkbruksmønstra på den nye staden Narvik.

## Litteratur

- Bauer, Laurie. 2008. A question of identity: A response to Trudgill. *Language in Society* 37(2), 270–73.
- Bjørkum, Andreas. 1968. *Årdalsmål hjå eldre og yngre: ei utgreiing om formverket med eit tillegg om generasjonsskilnad i lydverk, ordbruk og formverk*. Oslo: Universitetsforlaget.
- Bjørkum, Andreas. 1974. *Generasjonsskilnad i indresognsmål: talemålet i industribygdi Årdal og dei fire næraste jordbruksbygdene*. Oslo: Universitetsforlaget.
- Britain, David. 2012. Koineization and cake baking: Reflections on methods in dialect contact research. I: *Methods in Contemporary Linguistics*, red. Andrea Ender, Adrian Leemann og Bernhard Wälchli, 219–37. Berlin: Mouton de Gruyter.
- Coupland, Nikolas. 2008. The delicate constitution of identity in face-to-face accommodation: A response to Trudgill. *Language in Society* 37(2), 267–70.
- Det statistiske Centralbureau. 1902. *Folketellingen i Kongeriget Norge. 3 December 1900. Første Hefte. Folkemængde i Rigets forskjellige administrative Inddelinger m. v.* Kristiania: Aschehoug. Digitalt tilgjengeleg frå <https://www.ssb.no/a/folketellinger/>
- Det statistiske Centralbureau. 1903. *Folketellingen i Kongeriget Norge. 3 December 1900. Andet Hefte. Folkemængde fordelt efter Kjønn, Alder og ækteskabelig Stilling*. Kristiania: Aschehoug. Digitalt tilgjengeleg frå <https://www.ssb.no/a/folketellinger/>
- Det statistiske Centralbyraa. 1912. *Folketellingen i Norge. 1 december 1910. Første hefte. Folkemængde i rikets forskjellige administrative inndelinger m. v.* Kristiania: Aschehoug. Digitalt tilgjengeleg frå <https://www.ssb.no/a/folketellinger/>
- Digitalarkivet. 2000 [1900]. *Folketelling 1900 for 1855 Ankenes herred*. Oslo: Digitalarkivet. Digitalt tilgjengeleg frå: [www.digitalarkivet.no](http://www.digitalarkivet.no)



- Frank, Gustaf. 1950. *Kiruna, 1900–1950; Minnesskrift till 50-års-jubileet*. Stockholm: Esselte reklam.
- Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret MacLagan, Andrea Sudbury og Peter Trudgill. 2004. *New Zealand English: its origins and evolution*. Cambridge: Cambridge University Press.
- Hagen, Willy. 1953. *Innflyttingen til industristedet Sauda. Vandringsveier og vandringsmåter*. Hovudfagsoppgåve, Universitetet i Oslo.
- Hernes, Reidunn. 2006. *Talemål i endring? Ein longitudinell studie av talemålsutvikling og språkleg røyndomsoppfatning hjå ungdommar i Os*. Doktoravhandling, Universitetet i Bergen.
- Hickey, Raymond. 2003. How do dialects get the features they have? On the process of new dialect formation. I: *Motives for Language Change*, red. Raymond Hickey, 213–40. Cambridge: Cambridge University Press.
- Hoel, Oddmund Løkensgard. 2018. Unionstida med Sverige (1814–1905). I: *Norsk språkhistorie. Bind 4. Tidslinjer*, red. Agnete Nesse, 425–502. Oslo: Novus.
- Hoff, Ingeborg. 1960. Norsk dialektatlas. *Norsk tidsskrift for sprogvidenskap XIX*, 595–622.
- Holmes, Janet og Paul Kerswill 2008. Contact is not enough: A response to Trudgill. *Language in Society* 37(2), 273–77.
- Jahr, Ernst Håkon. 2007. Bruk av omgrepa 'standardtalemål', 'normalisering' og 'knot' for å skildre språktilhøva i Noreg i dag. I: *Å sjå samfunnet gjennom språket. Heidersskrift til Helge Sandøy på 60-årsdagen 14.06.2007*, red. Gunnstein Akselberg og Johan Myking, 93–8. Oslo: Novus forlag.
- Jahr, Ernst Håkon. 2008 [1984]. Synet på bydialektene i forrige hundreår og frem til ca. 1910. I: *Språkhistorie og språkkontakt. Language History and Language Contact. Festskrift til Ernst Håkon Jahr på 60-årsdagen 4. mars 2008*, red. Geir Wiggen, Tove Bull og Marit Aamodt Nielsen, 109–13. Oslo: Novus.
- Jahr, Ernst Håkon. 2013. Dialect ecology: the case of Norway – history and background. I: *Language Ecology for the 21<sup>st</sup> Century: Linguistic Conflicts and Social Environments*, red. Wim Vandenburg, Ernst Håkon Jahr og Peter Trudgill, 189–205. Oslo: Novus.
- Kerswill, Paul. 1994a. Dialektkontakt og sosiolingvistiske strukturer i Norge og i England. I: *Dialektkontakt, språkkontakt och språkförändring i Norden*, red. Ulla-Britt Kotsinas og John Helgander, 220–31.

Stockholm: Meddelanden från institutionen för nordiska språk vid Stockholms universitet.

Kerswill, Paul. 1994b. *Dialects converging: rural speech in urban Norway*. Oxford: Clarendon Press.

Kerswill, Paul. 2007. New-dialect formation: The inevitability of colonial Englishes (review). *Language* 83(3), 657–61.

Kerswill, Paul. 2010. Contact and New Varieties. I: *The Handbook of Language Contact*, red. Raymond Hickey, 230–51. Chichester: Wiley-Blackwell.

Kerswill, Paul. 2013. Koineization. I: *Handbook of Language Variation and Change*, 2.utg., red. J. K. Chambers og Natalie Schilling, 519–36. Oxford: Wiley-Blackwell.

Kerswill, Paul og Peter Trudgill. 2005. The birth of new dialects. I: *Dialect Change: Convergence and Divergence in European Languages*, red. Paul Kerswill, Peter Auer og Frans Hinskens, 196–220. Cambridge: Cambridge University Press.

Kerswill, Paul og Ann Williams. 2000. Creating a new town koine: Children and language change in Milton Keynes. *Language in Society* 29, 65–115.

Kleivenes, Kristine Tokle. 2002. *Med kristiansundsdiakten på flyttefot: en undersøkelse av seks utflytta kristiansunderes språkholdninger, identifikasjoner og språklige atferd i det "nye" språkklimaets tid*. Hovudfagsoppgåve, Universitetet i Bergen.

Labov, William. 1994. *Principles of Linguistic Change. Internal factors*. Oxford: Blackwell.

Labov, William, Sharon Ash og Charles Boberg. 2006. *The Atlas of North American English*. Berlin/New York: Mouton de Gruyter.

Mæhlum, Brit. 1986. *Språklige variasjonsmønstre hos innflyttere i Oslo*. Oslo: Novus.

Mufwene, Salikoko. 2001. *The Ecology of Language Evolution*. Cambridge: Cambridge University Press.

Mufwene, Salikoko. 2006. The comparability of new-dialect formation and creole development. *World Englishes* 25(1), 177–86.

Mufwene, Salikoko. 2008. Colonization, population contacts, and the emergence of new language varieties: A response to Peter Trudgill. *Language in Society* 37(2), 254–59.

Nesse, Agnete. 2008. *Bydialekt, riksmål og identitet: sett fra Bodø*. Oslo: Novus forlag.

- Nesse, Agnete. 2015. Bruk av dialekt og standardtalemål i offentligheten i Norge etter 1800. I: *Talemål etter 1800. Norsk i jamføring med andre nordiske språk*, red. Helge Sandøy, 89–111. Oslo: Novus.
- Neteland, Randi. 2009. «vann og kan og mann» *En sosiolingvistisk studie av nn-variabelen og andre språktrekk blant ungdommer i Narvik*. Masteroppgåve, Universitetet i Bergen.
- Neteland, Randi. 2013. Industristedtalemål og dialektendring. *Norsk Lingvistisk Tidsskrift* 1, 76–97.
- Neteland, Randi. 2014. Innflyttertalemåls innvirkning på talemålsendringer i industristeder. *Maal og minne* 106(1), 1–66.
- Neteland, Randi. 2017. Koine formation in context. *Journal of Historical Sociolinguistics* 3(1), 37–54.
- Neteland, Randi. 2019. *Koine formation and society*. Lanham: Lexington Books.
- Omdal, Helge. 1979. Høyangermålet – en ny dialekt. I: *Språklig samling på folkemåls grunn. Artikler frå bladet Språklig Samling*, red. Lars S. Vikør og Geir Wiggen, 75–7. Oslo: Novus.
- Omdal, Helge. 1994. *Med språket på flyttefot: språkvariasjon og språkstrategier blant setesdøler i Kristiansand*. Doktorgradsavhandling, Uppsala universitet.
- Orton, Harold, Stewart Sanderson og John Widdowson. 1978. *The Linguistic Atlas of England*. London: Croom Helm.
- Rambø, Gro-Renée. 2018. Det selvstendige Norge (1905–1945). I: *Norsk språkhistorie. Bind 4. Tidslinjer*, red. Agnete Nesse, 503–602. Oslo: Novus.
- Sandve, Bjørn Harald. 1976. *Om talemålet i industristadene Odda og Tysedal: generasjonsskilnad og tilnærming mellom dei to målføra*. Hovedfagsoppgåve, Universitetet i Bergen.
- Sandøy, Helge. 1985. *Norsk dialektkunnskap*. Oslo: Novus.
- Sandøy, Helge. 2000. Utviklingslinjer i moderne norske dialektar. *Folkemålsstudier* 39, 345–84.
- Sandøy, Helge. 2011. Language culture in Norway. A tradition of questioning standard language norms. I: *Standard Languages and Language Standards in a Changing Europe*, red. Tore Kristiansen og Nikolas Coupland, 119–26. Oslo: Novus.
- Sandøy, Helge og Agnete Nesse. 2016. Språkendring. I: *Norsk språkhistorie. Bind 1. Mønster*, red. Helge Sandøy, 31–99. Oslo: Novus.

- Schreier, Daniel. 2014. On cafeterias and new dialects. The role of primary transmitters. I: *The Evolution of Englishes. The Dynamic Model and beyond*, red. Sarah Buschfeld, Thomas Hoffmann, Magnus Huber og Alexander Kautzsch, 231–48. Amsterdam: John Benjamins.
- Schneider, Edgar W. 2008. Accommodation versus identity? A response to Trudgill. *Language in Society* 37(2), 262–67.
- Siegel, Jeff. 2001. Koine formation and creole genesis. I: *Creolization and Contact*, red. Norval Smith og Tonjes Veenstra, 175–97. Philadelphia: John Benjamins Publishing Company.
- Simonsen, Hanne Gram, Kristian Emil Kristoffersen, Dorthé Bleses, Sonja Wehberg og Rune N. Jørgensen. 2013. The Norwegian Communicative Development Inventories: Reliability, main developmental trends and gender differences. *First language* 1, 1–21.
- Solheim, Randi. 2006. *Språket i smeltegryta: sosiolingvistiske utviklingslinjer i industrisamfunnet Høyanger*. Doktorgradsavhandling, Universitetet i Trondheim.
- Solheim, Randi. 2009. Dialect development in the melting pot: The formation of a new culture and a new dialect in the industrial town of Høyanger. *Nordic Journal of Linguistics* 32(2), 191–206.
- Statistisk Sentralbyrå. 2021. Norske folketeljingar. <https://www.ssb.no/a/folketellinger/> Sist besøkt 13.07.2021.
- Språksamlingane. 2021. *Målføresynopsisen*. <https://www.uib.no/ub/114703/målføre-og-dialektologi> og <http://inger.uib.no/perl/search/search.cgi?appid=145&tabid=2165> Sist besøkt 13.07.2021. [Ny side for søk: <https://synopsis.uib.no>]
- Storm, Johan. 1884. Kortere Ordliste med Forklaring af Lydskriften. [Kristiania]: [ukjent].
- Sørli, Mikjel. 1959. Om Odda-målet i dag. *Maal og Minne* 1959, 55–66.
- Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Blackwell.
- Trudgill, Peter. 2004. *New-dialect formation. On the Inevitability of Colonial Englishes*. Edinburgh: Edinburgh University Press.
- Trudgill, Peter. 2008a. Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society* 37(2), 241–80.
- Trudgill, Peter. 2008b. On the role of children, and the mechanical view: A rejoinder. *Language in Society* 37(2), 277–80.

- Tuten, Donald. 2008. Identity formation and accommodation: Sequential and simultaneous relations. *Language in Society* 37(2), 259–62.
- Utter, Harald. 2010. *Å leve med fabrikkens puls - En studie av hagebyen Åbøbyen i Sauda og industriarbeiderne som bodde der i perioden 1960 til 1980*. Doktorgradsavhandling, Universitetet i Oslo.
- Wetås, Åse. 2014. Nordisk dialektkorpus og Norsk målføresynopsis – ei samanlikning av to kjelder til talemålsforskning frå kvar si tid. I: *Språk i Norge og nabolanda. Ny forskning om talespråk*, red. Janne B. Johannessen og Kristin Hagen, 315–35. Oslo: Novus.
- Aarhun, Tronn. 1973. *Befolkningsutvikling og endringer av den yrkesmessige struktur i Sauda 1900-1920*. Hovedfagsoppgave, Universitetet i Bergen.
- Aas, Steinar. 2001. *Byen, banen og bolaget. Narviks historie, bind 1, 1902–1950*. Fagernes: Stiftelsen Narviks Historieverk.

## Abstract

The paper's point of departure is a theoretical question regarding the factors that play a part in the koine formation process, namely if there is a correlation between the dialect background of in-migrants to a new town and the linguistic features used in the emerging new dialect. This question is tested using a combination of the population census from Narvik in 1900 and the dialect synopsis ("målføresynopsisen") on traditional Norwegian dialects. While many previous studies on industrial town dialects and similar new town dialects have concluded that the in-migrants' dialects have a significant influence on the outcome of the new dialect formation, this paper shows that the dialects the in-migrants brought with them from their hometowns, did not have a direct influence on how the feature investigated here, is pronounced in the new dialect.

Randi Neteland  
 Universitetet i Bergen  
 Randi.Neteland@uib.no





# Fra gammelt seddelarkiv til nye forskningsresultater: Nytt lys på gammelnorsk

Av Juliane Tiemann

Denne artikkelen tar for seg livssyklusen til datamateriale med utgangspunkt i det fysiske materialet i Språksamlingene ved Universitetet i Bergen. Materialet er gamle ordbokssedler over norrønt/gammelnorsk som senere ble overført til digitale løsninger. Det blir skissert hvordan dette materialet har blitt brukt om igjen for å bygge opp og supplere et nytt digitalt korpus (Ko-NoKs = Korpus over den norske *Konungs skuggsjá*) utformet for å besvare nye forskningsspørsmål. Som eksempel på en ny analyse basert på dette gamle forskningsmaterialet, rapporterer artikkelen om noen resultater fra en korpusstudie av syntaktisk variasjon i gammelnorsk.

## 1 Introduksjon

Studier av eldre språk er i sin natur korpusbaserte. Analysearbeid krever først og fremst at materialet er tilgjengelig, helst allerede i digital form, og at korpus inneholder opplysninger som trengs for analysen. I dag kan vi bygge videre på allerede eksisterende korpus for språkmaterialet til eldre norsk og norrønt i både fysisk og digital form. Språksamlingene ved Universitetet i Bergen inneholder blant annet et seddelarkiv (opprettet under arbeidet med det tidligere Gammalnorsk ordboksverk<sup>1</sup>) med om lag 650 000 ordbokssedler over norrønt/gammelnorsk. En god del av denne ressursen ble overført til digitale arkiver under Dokumentasjonsprosjektet ved Universitetet i Oslo på 1990-tallet. Men opplysningene for flere tusen ord er fremdeles kun lagret på sedler som ligger i arkivskuffene i Språksamlingenes depot.

1. Del av Dokumentasjonsprosjektet (delprosjekt gammelnorsk) som var et samarbeid mellom de humanistiske miljøene ved Universitetet i Bergen, Universitetet i Oslo, Universitetet i Trondheim og Universitetet i Tromsø.

Denne artikkelen beskriver de forskjellige stegene fra det fysiske materialet og merking av sedlene i seddelarkivet til oppretting av digitale ressurser og analyse av ordstillingsvariasjon i gammelnorsk, basert på et nytt XML-basert korpus. Ved å bruke det fysiske materialet i Språksamlingene bygges det videre på det språkdokumentariske arbeidet som ble satt i gang for nesten 100 år siden. Rammen for arbeidet er PhD-prosjektet «Syntactic Variation and Information Structure in Old Norwegian: An Investigation of *Konungs skuggsjá* in AM 243 bā fol.» ved Universitetet i Bergen. Dette prosjektet fokuserer særlig på samspill mellom syntaks og informasjonsstruktur. Informasjonsstruktur er et komplekst fenomen som refererer til grensesnittet mellom struktur og betydning av språklige ytringer, og som opererer både på setnings- og kontekstnivå. For en analyse av informasjonsstrukturelle virkninger på den endelige setningsstrukturen kreves det derfor en detaljert analyse av språkmaterialet med merking på forskjellige lingvistiske nivåer. Dette er et stort og tidkrevende arbeid, der analysen som allerede har blitt gjort på ordbokssedlene, er en ressurs som forenkler denne prosessen, og som i tillegg forhindrer at det samme arbeidet må gjøres flere ganger. På grunnlag av eksisterende seddelmateriale ble det nye korpuset (KoNoKs = Korpus over den norske *Konungs skuggsjá*) utvidet med annotasjon for flere språklige detaljer spesifikt knyttet til analyse spørsmålene i studien (se kap. 3). I artikkelen vil et forskningseksempel brukes for å vise fram både arbeidet med det gammelnorske materialet og hvilke typer resultater som kan komme ut av et slikt forskningsprosjekt (se kap. 4). Før beskrivelsen av arbeidet med seddelarkivet vil den teoretiske rammen for dette arbeidet skisseres i kap. 2. Det vil også legges vekt på betydningen av analyser av gammelnorsk språkmateriale separat fra gammelislandsk språkmateriale.

## 2 Syntaktisk variasjon: Nytt prosjekt om gammelnorsk

Variasjon i og endring av ordstilling over tid er godt dokumentert for eldre germanske språk. I løpet av de siste tiårene har fokuset i språkstudiene skiftet mot analyser og beskrivelser av de mekanismene som styrer og utløser variasjon, der forskere særlig har sett på informasjonsstruktur og prosodisk vekt (jf. Taylor og Pintzuk 2012 a,b; Elenbaas og van Kemenade 2014; Walkden 2014). Selv om dette er et aktuelt forskningstema



som er hyppig diskutert for flere språk, mellom annet gammelengelsk, gammelhøytysk og gammelislandsk, har syntaktisk variasjon i gammelnorsk ikke fått mye oppmerksomhet i studier innen vestnordisk. Beskrivelser av syntaktisk variasjon i den vestnordiske språkgreinen skiller ikke mellom gammelislandsk (ca. 1100–1540) og gammelnorsk (ca. 1050–1350), og det blir antatt at det som gjelder for gammelislandsk også gjelder for gammelnorsk. Dette skjer selv om studier innen syntaks hovedsakelig har blitt utført på grunnlag av gammelislandsk materiale (jf. t.d. Thorbjörg Hróarsdóttir 2000, 2009). Praksisen med å ikke skille mellom disse to språkene på syntaktisk nivå blir gjentatt i anerkjent oversikt litteratur over norrønt språk, der t.d. Faarlund (2004: 2) skriver at “[t]here are a few minor phonological and morphological differences between Old Icelandic and Old Norwegian, but no known syntactic differences”. Også Mørck (2016: 320f.) viderefører dette i den nye norske språkhistorien, *Norsk språkhistorie I: Mønster*, der han skriver at «[s]iden det ser ut til å være ganske små syntaktiske forskjeller innanfor det norrøne området i denne perioden, kan vi tillate oss å gå ut fra at beskrivelsene av norrønt med sitt i stor grad islandske grunnlag i det store og heile er gyldige også for gammelnorsk.»<sup>2</sup>

Hovedgrunnen til at kildene for forskjellige typer analyser primært er islandske og ikke norske, er at omfanget på kilder og opparbeidet/annotert korpusmateriale er større for gammelislandsk (jf. The Icelandic Parsed Historical Corpus/IcePaHC)<sup>3</sup> enn for gammelnorsk. Både synkrone og diakrone analyser av gammelnorsk syntaks gir dermed et ufullstendig bilde, mens ordstillingsvariasjon i gammelislandsk har blitt analysert fra flere forskjellige vinkler og med materiale som strekker seg over flere år-

2. Slike påstander er basert på ordstillingsvariasjon som kan observeres i begge språk, men tar ikke inn detaljerte komparative analyser. Slike analyser kan imidlertid vise tendenser for forskjellige ordstillingsmønstre, og med dette også syntaktiske forskjeller mellom disse språkene. Ulike tendenser i språkene adskiller disse over lengre tid som synliggjør forskjeller mellom islandsk og norsk (og islandsk og de andre fastlandsskandinaviske språkene), t.d. er rekkefølgen av det finitte verbet og setningsadverbialer i leddsetninger forskjellig. I moderne islandsk står verbet i andre posisjon, både i hoved- og leddsetninger (jf. Franco 2010: 26f.; Holmberg og Platzack 2008: 437f.), mens i moderne norsk følger verbet setningsadverbialet. Også rekkefølgen av PP-er relativt til NP-er viser forskjeller mellom insulære og fastlandsskandinaviske språk (jf. Thorbjörg Hróarsdóttir 2000: 191).
3. [http://www.linguist.is/icelandic\\_treebank/](http://www.linguist.is/icelandic_treebank/). Versjon 0.9 (2011). Prosjektet der korpuset ble bygget opp, ble finansiert av Icelandic Research Fund, U.S. National Science Foundation og Rannsóknasjóður Háskóla Íslands. Se Wallenberg m.fl. (2011).

hundrer (jf. bl.a. Halldor Sigurðsson 1985, 1988, 1994; Thorsteinn Indriðason 1987; Eiríkur Rögnvaldsson 1994/1995, 1996; Thorbjörg Hróarsdóttir 1996, 2000, 2009; Haugan 2000).<sup>4</sup> Fraværet av gammelnorsk i slike studier er en god grunn til å fokusere mer på dette språklige materialet. Behovet for syntaktisk annoterte gammelnorske tekster ble først fanget opp i Menotec-prosjektet<sup>5</sup> som er publisert i trebanken til PROIEL.<sup>6</sup> I dette prosjektet valgte man en dependensanalyse som syntaktisk modell. Bortsett fra KoNoKs-korpuset finnes det så langt ingen gammelnorske korpus som er bygget opp etter frase-strukturelle analyser. KoNoKs ble utviklet i tilknytning til forskningsprosjektet «Syntactic Variation and Information Structure in Old Norwegian» som gir en analyse av ordstillingsvariasjon i gammelnorsk fra et pragmatisk perspektiv med søkelys på påvirkningen av informasjonsstruktur og av prosodisk vekt som avgjørende faktorer for ulik ordstilling. Fremgangsmåten for dette var en næranalyse av én gammelnorsk tekst (se kap. 3) med fokus på blant annet variasjon i objektsposisjon, adjektivposisjon og possessivkonstruksjoner. Dataene fra analysene ble sammenlignet med resultater fra lignende analyser for gammelisländsk (særlig Thorbjörg Hróarsdóttir 2000, 2009; Pfaff 2019). Et slikt komparativt perspektiv med gammelisländsk danner et bedre grunnlag for enten å bekrefte at beskrivelser for gammelisländsk syntaks også er gyldige for gammelnorsk, eller for å vise at det faktisk finnes forskjeller mellom de to språkene (og dermed eksisterende forskningshull). Korpuset KoNoKs er dermed et nytt, nyttig og viktig grunnlag for videre arbeid med og analyser av syntaktisk variasjon i gammelnorsk og for videre komparative studier, både når det gjelder syntaks og informasjonsstruktur. Noen av resultatene for syntaktisk variasjon med fokus på objektsposisjonen, dvs. opposisjonen i ordstillingene OV og VO og hvordan informasjonsstruktur og prosodisk vekt styrer denne variasjonen, vil kort skisseres i denne artikkelen.

4. Selv om vi har en rekke publikasjoner som fokuserer på gammelnorske kilder fra forskjellige perspektiver (jf. Flom 1921–1923; Holm-Olsen 1952; Schnell og Simek 2000; Christoffersen 1994, 2002; Haugen og Ommundsen 2010), så er disse studiene ofte mer begrensede og det finnes en mindre mengde lingvistiske analyser enn for gammelisländsk.
5. <https://www.menota.org/menotec.xml>. Dette er et infrastrukturprosjekt finansiert av Norges forskningsråd (2010-2012).
6. <https://proiel.github.io/>. Den syntaktiske analysen i dette prosjektet er basert på et dependens-grammatisk rammeverk og skiller seg dermed fra analysen gjort i PhD-prosjektet som er presentert her.

2.1 *Det teoretiske rammeverket: Kort om informasjonsstruktur*

Ifølge Lambrecht (1994: xv) kan informasjonsstruktur bli forstått som en grammatisk komponent som styrer «the relationship between speaker assumptions and the formal sentence structure». Han (1994: xiii) skriver videre at

[i]n the information structure component of language, propositions as conceptual representations of states of affairs undergo pragmatic structuring according to the utterance contexts in which these states of affairs are to be communicated. Such pragmatically structured propositions are then expressed as formal objects with morphosyntactic and prosodic structure.

Hinterhölzl og van Kemenade (2012) antar at ordstilling er en del av setningsgrammatikken, mens informasjonsstruktur er en del av pragmatikken. På denne måten styrer syntaksen hvordan en setning kan se ut (her refereres det til alle mulige overflatestrukturer samt restriksjoner i et språk), mens pragmatikken bestemmer den spesifikke formen av en setning som passer i en gitt kontekst. Hinterhölzl og van Kemenade (2012: 819) argumenterer videre for at «syntax, [information structure], and prosody all three influence the choice for a particular word order pattern» som viser at «[information structure] interacts with the entire grammatical system, rather than with just one component.» Informasjonsstruktur er altså et komplekst fenomen og definisjon og tilhørende kategorier er hyppig diskutert i litteraturen (jf. Levinson 1983; Kruijff og Duchier 2003; Kruijff-Korbayová og Steedman 2003). I denne artikkelen spiller denne diskusjonen og bruken av forskjellig terminologi (tema-remå vs. topik-fokus, osv.)<sup>7</sup> en mindre rolle. Her vil det informasjonsstrukturelle nivået, noe forenklet, defineres ved bruk av den følgende passasjen gitt i Matic (2015: 95):

Utterances transmit both the information contained in the message and the implicit or explicit instructions on how this information is to be processed and integrated into the hearer's knowledge stock. ... The information comes about by relating the 'new' elements (i.e., what the hearer is assumed not to be aware of) to these 'given' elements.

7. Dikotomien tema-remå (internasjonalt bruker man heller *topic-comment* som begreper der *comment* inneholder et framhevet fokuselement) beskriver ulike deler av en setning som viser til hva setningen dreier seg om (tema) og hva som er ny informasjon relatert til setningens tema (remå).

Dette beskriver altså hvordan informasjon er presentert i den aktuelle kommunikative utvekslingen (form, uttrykk og strukturering), der kontekstinformasjon/en diskurs fortløpende oppdateres med ny informasjon. Dette er et rekursivt fenomen der informasjonen deles opp i forståelige diskursenheter (jf. Chafe 1976 «information packaging») relatert til det som allerede er kjent i en kontekst (jf. Roberts 2012 «questions under discussion»). Som nevnt ovenfor opererer informasjonsstruktur både på setningsnivå og diskursnivå.

Det teoretiske rammeverket for analyse av informasjonsstruktur ble opprinnelig utviklet for analyse av moderne språk med sterkt fokus på intonasjonsmønstre. Noen av de mest fremtredende informasjonsstrukturelle kjennetegnene slik som prosodisk prominens (trykk) er ikke direkte tilgjengelige i historisk materiale. For en informasjonsstrukturell analyse av språklig materiale uten tilgang til et aktivt intonasjonssystem er det derfor viktig å kunne bygge på veldefinerte annotasjonskategorier på flere språklige nivåer og utprøvde analysemetoder. Flere arbeider de siste tiårene setter søkelys på nettopp dette, og inkluderer store mengder korpusmateriale for forskjellige eldre språk som latin, hettittisk, gammelholtytsk og gammelengelsk.<sup>8</sup>

8. Flere store prosjekter som tar for seg informasjonsstrukturelle analyser for både moderne og eldre språk er finansiert av Deutsche Forschungsgemeinschaft: a) SFB-prosjektet «Information structure: the linguistic means for structuring utterances, sentences and texts» (<http://www.sfb632.uni-potsdam.de/>, 2003-2015), b) «Informationsstruktur in älteren indogermanischen Sprachen» (<https://gepris.dfg.de/gepris/projekt/109055449?context=projekt&task=showDetail&id=109055449&>, 2009–2016) og c) «Informationsstruktur in komplexen Sätzen – synchron und diachron: Erarbeitung eines diachronen Kontrastkorpus» (<https://gepris.dfg.de/gepris/projekt/199843560>, 2011–2017). I Norge ble det gjennomført et prosjekt som fokuserer på germanske og romanske språk, finansiert av Norges forskningsråd og Institutt for litteratur, områdestudier og europeiske språk ved Universitetet i Oslo: «Information Structure and Word Order Change in Germanic and Romance Languages» (<https://www.hf.uio.no/ilos/english/research/projects/iswoc/index.html>, 2011–2014); jf. Bech og Eide 2014). Dette prosjektet inkluderte også en analyse av gammelnorsk, men selv om norsk var en del av dette prosjektet, ble kun deler av *Óláfs saga ins helga* (Upps DG 8 II) analysert under arbeidet, og så langt finnes det ingen publikasjoner som tar for seg resultatene for det gammelnorske materialet. I tillegg er det viktig å poengtere at annotasjonen i ISWOC følger en dependens-grammatisk analyse (i samme tråd som analysene presentert i PROIEL) som viser til hierarkisk ordning av konstituenten i motsetning til en frase-strukturell lineær analyse.

2.2 Annotasjonsarbeid: Kort om KoNoKs

For å kunne ekstrahere informasjonsstrukturelle entiteter (fokus og topik/tema) og for å kompensere for manglende intonasjonale faktorer i det eldre språklige materialet på en effektiv måte, må analysen bygge på en kumulativ oppsummering av tydelige kategorier i forskjellige strukturelle domener. En multi-modal korpustilnærming<sup>9</sup> er dermed nødvendig (jf. også Donhauser 2007; Petrova m.fl. 2009). Denne annotasjonen tar utgangspunkt i en omfattende grammatisk merking inklusive lemmatisering av materialet. For KoNoKs er den første delen av annotasjonen således det som ble gjort på ordbokssedlene; dette er grunnlaget for alt videre arbeid. Den videre annotasjonen følger koherente annotasjonsretningslinjer for eldre korpusspråk som har blitt skissert i nasjonale og internasjonale prosjekter (se fn. 8). Annotasjonen av informasjonsstrukturelle enheter bygger på en kombinasjon av særlig semantiske, pragmatiske og diskursrelaterte annotasjoner, slik som “saliency” (animathet), “givenness”, “definiteness” eller “anaphoric/cataphoric relations”. På den måten kan et element som er annotert for å være kjent, bestemt og har anaforisk referanse, analyseres som en sterk topik-kan-

	179	180	181	182	183	184	185	186	187	188
[text]	cn	þcir	isar	liggja	mæirr	i	landnorðr	eða	til	norðs
[lem]	but	this	ice	lie	more	in	north.land	or	to	north
[trans]	en	sá	íss	liggja	mæir	i	landnorðr	eða	til	norðr
[glos-morph]		NOM.M.PL	NOM(M).PL	PRS.IND.ACT.3PL	comp		ACC(N).SG			GEN(N).SG
[pos]	conj	prdem	noun	v.fin	adv	prep	noun	conj	prep	noun
[v-sem]										
[saliency]		pr3.dem	abstract	pr3			abstract			abstract
[givenness]		access-sit	giv				new			new
[definitness]		def					indef			indef
[gramm-def]			indef							
[context]		demon/dict.ref								
[frame]		schemeice								
[l-particle]										
[shift]		continue								
[Topic]		Con-T								
[position T]		initial								
[Comment]	cn			cm						

Fig. 1: Skjermdump av annotasjonen for en topik-kandidat (10r, kol.a:26–27).

9. Et multi-modalt korpus er en datakolleksjon som brukes for å studere hvordan to eller mer modaliteter (forskjellige nivåer) knyttes sammen og påvirker hverandre.

	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196
[text]	þeir	isar	liggja	meiðr i		larndnorðr	eða til	norðrs	firi	larndino	hæll	en til	suðrs	eða	utsuðrs		
[trans]	this	ice	lie	more in		north.land	or to	north	country	rather	but to	south	or	southwest			
[lem]	sá	iss	liggja	meiðr i		landnorðr	eða til	norðr	fyri	land	heldr	en til	suðr	eða	útsuðr		
[glos-morph]	NOM.M.PL	NOM(M).PL	PRS.IND.ACT.3PL	comp		ACC(N).SG		GEN(N).SG	DAT(N).SG	comp			GEN(N).SG		GEN(N).SG		
[pos]	prdem	noun	v.fin	adv	prep	noun	conj	prep	noun	prep	noun	adv	adv	prep	noun	conj	noun.prop
[v-sem]																	
[salience]	pr3.dem	abstract	pr3			abstract		abstract	abstract						abstract		abstract
[givenness]	access-sit	giv						new	giv						new		new
[definitness]	def					indef		indef	def						indef		indef
[gramm-def]		indef															
[context]	demon/dict.ref																
[frame]	scheme:ice																
[l-particle]										scheme:greenland							
[shift]	continue																
[Topic]	Con-T																
[position-T]	initial																
[Comment]			cm														
[NFocus]																	
[CFocus]						cf									cf		
[position-F]						middle									final		

Fig. 2: Skjermdump av annotasjonen for en fokus-kandidat (10r, kol.a:26–29).

didat og et element som er annotert som nytt, ubestemt og kontrastivt som en sterk fokus-kandidat. Den kumulative analysen som viser forskjellige og overlappende trekk, er illustrert i figur 1 og 2 (over).<sup>10</sup>

For slikt analysearbeid er det som nevnt nødvendig at språkmaterialet er opparbeidet med bred grammatisk informasjon (for en detaljert beskrivelse av annotasjonen se Tiemann upubl. manuskript: Kappa).

### 3 Datamateriale: Fra sedler til datafil

Utarbeidede forskningsressurser for det gammelnorske språket eksisterer i dag både i fysisk og digital form. En stor del av de digitale ressursene med morfologisk merking har utgangspunkt i seddelarkivet i Språksamlingene. De gamle ordbokssedlene er oppbevart i skuffer og lagret i Språksamlingenes depot (ett av arkivene/magasinene til Språksamlingene i Bergen) som vist i figur 3.

Analysearbeidet i PhD-prosjektet som er presentert her, tar hovedsakelig for seg én større og sammenhengende gammelnorsk tekst, *Konungs skuggsjá* i det norske hovedhåndskriftet, AM 243 bα fol. Verket er en pedagogisk-filosofisk tekst i form av en dialog mellom far og sønn,

10. Referansen gitt i parenteser leses på følgende måte: manuskriptside (recto og verso), kolonne (a og b), og linjenummer for eksempelet.



Fig. 3: Språksamlingenes depot – seddelarkivet.

skrevet på midten av 1200-tallet. I omfang er *Konungs skuggsjá* ingen liten tekst, og selv om det er noen lakuner i AM 243 bā fol., inneholder dette håndskriftet 63 910 ord. Teksten er del av Språksamlingene og har her fått full grammatisk oppmerking på ordbokssedler (se kap. 3.1). Disse sedlene er lagret i spesiallagede syrefrie bokser og er sortert alfabetisk, som vist i figur 4.



Fig. 4: Gammelnorske ordbokssedler – Konungs skuggsjá.

### 3.1 Digitalisering

Hver ordboksseddel i samlingen har seddelformat ca. 8 x 18 cm og inneholder 6–10 linjer med tekst, som vist i figur 5. Dette sikrer at hvert opp-

slagsord (leksem) er gitt i en best mulig kontekst. Alle oppslagsordene i hvert tekststykke har hver sin seddel.

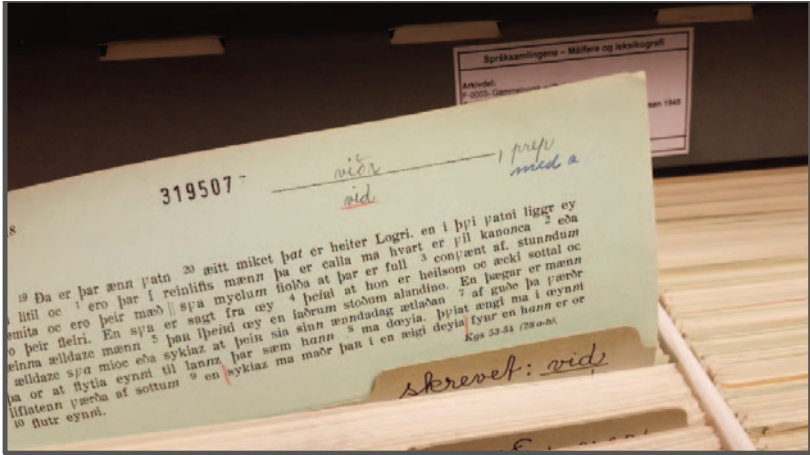


Fig. 5: Ordbokseddel – detalj.

Hvert ord har fått grammatisk bestemmelse (ordklasse og bøyningsform), og er ført opp som oppslagsord i normalisert form øverst på seddelen, ved siden av ordets skriveform i håndskriftet. Alt materialet som finnes på slike sedler, inklusive all grammatisk merking, burde vært overført til digitale databaser. Men selv om en stor del av Språksamlingene allerede er digitalisert, finnes det mellom annet i seddelarkivet fremdeles materiale som ennå ikke er fullstendig overført til digitale arkiver.

Middelaldertekstene fra samlingen blir innlemmet i Menota-arkivet<sup>11</sup> der også teksten til *Konungs skuggsjá* er tilgjengelig. Figur 6 viser en del av katalogen i Menota der Språksamlingene er nevnt som hovedredaktør. Som del av Dokumentasjonsprosjektet ble materialet på kartotek kortene fra den eldste fasen av det norske språket innlest, digitalisert og senere overført til menotisk XML,<sup>12</sup> og dermed gjort om til elektroniske tekster. Menotisk XML følger en TEI standard<sup>13</sup> og den tilføyde annotasjonen er på et morfologisk nivå i tråd med bestemmelse foretatt på ordboks-

11. <https://www.menota.org/forside.xhtml>, (2001 – fortløpende).  
 12. XML-kodingen som følger retningslinjene oppgitt for Menota. Se Menota Handbook 3.0: [https://www.menota.org/HB3\\_index.xml](https://www.menota.org/HB3_index.xml).  
 13. <https://tei-c.org/>.



## FRA GAMMELT SEDDELARKIV TIL NYE FORSKNINGSRISULTATER

Signature	MS or Work Title	Main Editor(s)	Date	Lang.	Words	Lemma	Form
AM 178 fol (1r.1-17r.17)	Piäriks saga af Bern (part 1 of 22)	Spräksamlingane	c. 1600-1700	isl	9216	party	party
AM 178 fol (47r.3-49r.9)	Piäriks saga af Bern (part 10 of 22)	Spräksamlingane	c. 1600-1700	isl	1375	completely	completely
AM 178 fol (58r.19-58v.6)	Piäriks saga af Bern (part 3 of 22)	Spräksamlingane	c. 1600-1700	isl	176	completely	completely
AM 178 fol (60r.16-61r.15)	Piäriks saga af Bern (part 5 of 22)	Spräksamlingane	c. 1600-1700	isl	672	completely	completely
AM 243 ba fol	Konungs skuggsjä	Spräksamlingane	c. 1275	nor	63910	completely	completely
AM 619 4to	Old Norwegian homily book	Spräksamlingane	c. 1200-1225	nor	62237	completely	completely
DG 4-7	Strengleikar	Spräksamlingane	c. 1270	nor	38453	completely	completely
Holm perg 17 4to	Thömass saga erkbiskups	Spräksamlingane	c. 1300	nor	59684	completely	completely

Fig. 6: Katalog i Menota.

sedlene. Målet med Dokumentasjonsprosjektet var å bygge opp et fullstendig ordarkiv med alle forekomster av ord i norsk språk fra de første skrevne beleggene på norsk fra omkring 1200. Dette skulle gi en samlet og enkel tilgang til norsk middelalderspråk.

Filene i Menota kan lastes ned, og har en omfattende 'header'<sup>14</sup> med utfyllende informasjon om blant annet annotatorer, endringer og lisens.

Figur 7 (neste side) viser hvordan samme tekststykke av *Konungs skuggsjä* ser ut i de forskjellige formatene fra faksimilefotografi til kartotekkort og til slutt i html-visning på Menotas hjemmeside, der teksten er publisert med en diplomatarisk transkripsjon (markeringen av ordet som er tatt opp på kortet, er tilføyd både i fotografi og i den elektroniske transkripsjonen).

Som nevnt ovenfor er en stor del av seddelarkivet allerede digitalisert og overført til Menota. Men omtrent en tredjepart av annotasjonen for *Konungs skuggsjä* hadde uheldigvis falt ut under konverteringen, og den grammatiske bestemmelsen for ordene i denne delen fantes utelukkende i seddelformat. Det var derfor nødvendig å gå tilbake til seddelarkivet for å fullføre annotasjonen som trengs for videre syntaktisk analyse av teksten. Det arkiverte fysiske materialet fikk i denne omgangen igjen en enorm verdi for forskning og kunne brukes som grunnlag til nye korpusanalyser. Uten arbeidet som skaperne av ordbokssedlene allerede hadde gjort, hadde det tatt så lang tid å fullføre all annotasjonen for prosjektet om syntaks og informasjonsstruktur i gammelnorsk at det sann-

14. En XML-header spesifiserer XML-versjonen og gir ytterligere informasjon om kodingen.

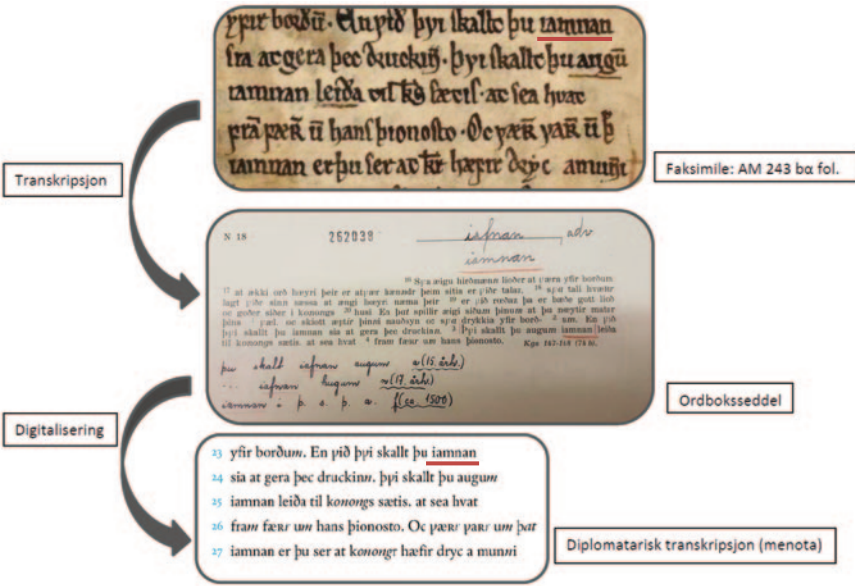


Fig. 7: Eksempel – tekststykke i de forskjellige formatene.

synligvis ville ha gått langt utover prosjektets omfang. Selv om mye av merkingen kan tilføyes nokså raskt i en XML-editor, slik at arbeidet generelt sett går mye forttere enn før det datamaskinelle arbeidet, er det fremdeles tidkrevende. Analysen som er skrevet på sedlene, er da en ressurs som forenkler arbeidsflyten med teksten og annotasjon (jf. også Tiemann 2018). I tillegg vil konsultering av dette materialet forhindre at det samme arbeidet må gjøres flere ganger.

### 3.2 Videre arbeid med datafilene

For den videre syntaktiske og informasjonsstrukturelle annotasjonen kunne arbeidet ikke lenger bli gjort i TEI XML-kodingen brukt i Menota siden det måtte defineres flere overlappende annotasjonsnivåer. I tillegg var det nødvendig å arbeide videre med teksten i en lineær visning, særlig for å se setningene bedre i kontekst og med mulighet for å ha både en horisontal og vertikal akse. For det videre arbeidet ble det derfor brukt en fritt tilgjengelig editor som også har blitt brukt innen store internasjonale forskningsprosjekter om informasjonsstruktur i eldre språk (jf.

fn. 8): EXMARaLDA-Partitur Editor.<sup>15</sup> Gjennom et XSLT-stilark<sup>16</sup> ble menotisk XML omdannet til EXMARaLDA-XML, slik at all morfologisk annotasjon samt selve teksten kunne vises i vertikale linjer som vist i figur 8. I tillegg ble det tilføyd flere annotasjonsnivåer med spesifisert verdi og noen med tom verdi i dette stilarket. Tomme nivåer kunne fylles inn ved behov under den videre analysen. EXMARaLDA gjør det mulig å definere og å sette annotasjonskategorier helt fritt. Som diskutert ovenfor er informasjonsstruktur et komplekst språklig fenomen som strekker seg over setningsgrenser og der konteksten spiller en viktig rolle. For å kunne peke på informasjonsstrukturelle kategorier er det derfor nødvendig å arbeide med flere overlappende nivåer samtidig.

Som vist i figur 8 er den horisontale akse brukt for verdiene av de forskjellige annotasjonene. Den vertikale akse viser de forskjellige annotasjonene der det er mulig å få en oppsummering av alle trekk i ulike strukturelle domener for å identifisere informasjonsstrukturelle komponenter (jf. fig. 1 og 2). Kategoriene for informasjonsstrukturelle kandidater er analysert helt mot slutten av annotasjonsarbeidet. På grunn av en lett tilgjengelig oppsummering blir en ekstraksjon av disse elementene transparent og enkel å komme tilbake til på et senere tidspunkt.

26	27	28	29	20	21	22	23	24	25	26	27	28
oc	heita	þau	iafnan	siðan	hins	hælgja	Kerinus	æpli	<p>	oc	fara	þau
ok	heita	ei	jafnan	eiðan	tunn	heilagr	Kerinus	æpli		ok	fara	ei
and	call	this	always	since	the	holy	Kerinus	apple		and	go	they
	PRS.IND.ACT.3PL	NOM.N.PL	pos	pos	OEN.M.SG	OEN.M.SG.pos	OEN(M).SG	NOM(N).PL			PRS.IND.ACT.3PL	NOM.N.PL
conj	v.ñn	prdem	adv	adv	prdem	adj	noun.prop	noun		conj	v.ñn	prdem
#												performative
pr3		pr3			pr3.dam		human	concrete			pr3	pr3
		access-rit			access-rit		grv	grv				access-rit
		def			def	def	def	indef				def
hactive		usa.ref			demon/defct.ref						hactive	usa.ref
		schema:apples					schema:Kerinus:apple					schema:apples
		chuff										continue
		S-T										Con-T
		middle										2P
om			om							om		
					rf							
					imal							

Fig. 8: EXMARaLDA-editor visning.

15. <https://exmaralda.org/en/partitur-editor-en/>.

16. Et XSLT-stilark beskriver gjennom et sett med regler en transformasjon fra et XML-dokument til et nytt XML-dokument.

Etter annotasjonen ble filene lastet opp i visualiserings- og søkeverktøyet ANNIS (ANNotation of Information Structure)<sup>17</sup> som inneholder en søkemonitor for å ekstrahere resultatene for den endelige analysen av materialet.

#### 4 Studie om gammelnorsk ordstilling

For å vise noen av resultatene som et slikt videre arbeid med gammelt materiale kan gi, vil siste del av denne artikkelen kort skissere en studie av syntaktisk variasjon og informasjonsstruktur i gammelnorsk som ble gjennomført i forbindelse med PhD-prosjektet «Syntactic Variation and Information Structure in Old Norwegian». Etter gjennomført ekstraksjon av dataer fra ANNIS kunne disse dataene analyseres innenfor en informasjonsstrukturell ramme og i tråd med modeller innen generativ syntaks. Det vil ikke bli gjort detaljert rede for selve analysen eller litteraturen i dette avsnittet. Se Tiemann (2022) for en fullstendig diskusjon.

##### 4.1 Objektsposisjon og informasjonsstatus

Studien av objektsposisjon følger særlig analysearbeider for gammelisländsk, der Thorbjörg Hróarsdóttir (2000, 2009) har undersøkt en stor mengde materiale fra et pragmatisk perspektiv. Tabell 1, hentet fra Thorbjörg Hróarsdóttir (2009), viser samspillet mellom informasjonsstatusen

Texts	New information			Old information		
	OV	VO	% OV	OV	VO	% OV
14th c.	75	204	26.9	156	36	81.3
15th c.	48	136	26.1	110	24	82.1
16th c.	50	132	27.5	77	17	82.0
17th c.	79	314	20.1	215	104	67.4
18th c.	20	98	16.9	39	43	47.6
19th c.	84	1,312	6.0	261	1,072	20.0

Tab. 1: Datasett for OV/VO og informasjonsstatus, Thorbjörg Hróarsdóttir (2009: 264).

17. <https://corpus-tools.org/annis/>. ANNIS ble designet i SFB-prosjektet nevnt ovenfor (se fn. 8).

(kjent – nytt)<sup>18</sup> og ordstilling av objekter (pre- eller postverbal stilling; relativt til hovedverbet) i islandsk.

Tabellen viser at objekter som bærer kjent informasjon, oftere står i en preverbal posisjon. Informasjonsstatusen til et objekt spiller dermed en viktig rolle for dets plassering i setningen. Dette følger den generelle og velkjente tendensen for grammatisk ordning der kjent informasjon kommer før ny informasjon. Siden forskningstradisjonen ikke skiller mellom gammelnorsk og gammelislandsk på et syntaktisk nivå, er nullhypotesen at man finner de samme resultatene for gammelnorsk. Det betyr at fordelingen av objekter<sup>19</sup> på OV/VO i gammelnorsk korrelerer med informasjonsstatusen til objekter på samme måte og i samme grad som i gammelislandsk. Vi forventer altså et relativt homogent sett

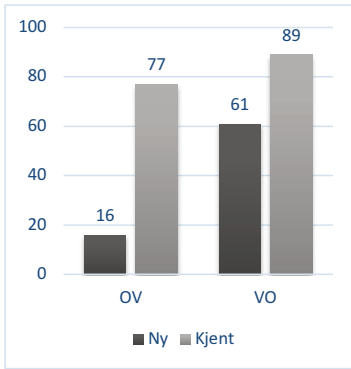


Fig. 9: Plassering av kjente og nye objekter i KoNoKs-datasettet.

av objekter i OV-stilling (kjent; beveger seg til en preverbal stilling på grunn av en informasjonsstrukturell utløser) og et heterogent sett av objekter i VO-stilling (ikke alle objekter vil nødvendigvis bli påvirket av utløseren til bevegelsen; andre faktorer kan forhindre bevegelsen). Datasettet tatt fra KoNoKs består av 243 setninger som har referensielle objekter i leddsetninger med to verb (dvs. hovedverb og hjelpeverb).<sup>20</sup> Figur 9 viser plasseringen av kjente og nye objekter i disse setningene (se Tiemann 2022 for gammelnorske eksempler).

18. Kjent informasjon er enten allerede nevnt i konteksten eller identifiserbar (dette inkluderer delt kunnskap og kunnskap om verden); ny informasjon er ikke nevnt før, uforutsigbar og ikke konkluderbar.
19. *Objekt* defineres her som komplement som er krevet av hovedverbet. Både nominale objekter og preposisjonalobjekter er tatt med i analysen.
20. Strukturer med et referensielt objekt og to verb sikrer at man har strukturer som viser bevegelse av objektet og ikke bevegelse av verbet. Setninger med negative og kvantifiserende objekter ble ekskludert fra studien. Disse objektene er ikke referensielle og kan vise ulik syntaktisk oppførsel (se Struik og van Kemenade 2018). I tillegg ble eksempler med uklar informasjonsstatus ekskludert (som i tilfeller av generisk referanse).

Resultatene ligner på det som Thorbjörg Hróarsdóttir (2009) har beskrevet for gammelisländsk med hensyn til objekters informasjonsstatus. Dette indikerer at behandlingen av gammelisländsk og gammelnorsk under fellesbegrepet 'norrønt' er legitimt når det gjelder det syntaktiske nivået. Men resultatene viser kun til en generell observasjon og til at man finner de samme ordstillingsmønstrene på overflatestrukturen med en lignende fordeling. En mer detaljert komparativ analyse av materialet og tallene (fra studien presentert her og fra Thorbjörg Hróarsdóttirs studie) tyder derimot på at de to språkene divergerer også på et syntaktisk nivå allerede på den tidsperioden som er undersøkt her. Resultatene av denne sammenligningen er oppsummert i tabell 2. Tallene i denne tabellen er sett i relasjon til VO-stillinger med nye og kjente objekter (tallene for VO-stillingen er ikke oppgitt i denne tabellen). For gammelisländsk er det oppgitt to århundrer (hentet fra Thorbjörg Hróarsdóttir 2009: 264) for å i tillegg vise en interessant utvikling med tanke på hvordan informasjonsstatus korrelerer med ordstilling over tid.

	OV-ny	OV-kjent
Gammelnorsk 13. årh.	20,8 % (16/77)	46,4 % (77/166)
Gammelisländsk 14. årh.	26,9 % (75/279)	81,3 % (156/192)
Gammelisländsk 18. årh.	16,9 % (20/118)	47,6 % (39/82)

Tab. 2: Prosent for OV-stilling i gammelnorsk og gammelisländsk.

Det første som er påfallende, er en nedgang av OV-stilling i gammelisländsk over tid. Også for norsk kan vi stadfeste at OV-stillinger forsvinner som mulige overflatestrukturer på et eller annet tidspunkt.<sup>21</sup> Men uten komplekse annoterte korpus for norsk for forskjellige århundrer er det vanskelig å diskutere dette i ytterligere detalj. Mer generelt viser tallene at påvirkningen av informasjonsstatusen er svekket allerede ganske tidlig i gammelnorsk sammenlignet med gammelisländsk. Sannsynligheten for at informasjonsstatus er en faktor for OV-stilling i gammelnorsk på 1200-tallet, er dobbelt så høy når objektet er kjent fra konteksten. For gammelisländsk er sannsynligheten tredoblet når objektet er kjent, og denne effekten er stabil over tid, selv om man finner en generell nedgang

21. Når man ser bort ifra topikalisering og 'stylistic fronting' og bare regner med den dominante ordstillingen som klassifiserer dette språket som et VO-språk. Se også 'Object Shift' i moderne isländsk (jf. Thorbjörg Hróarsdóttir 2009: 266, spes. 274ff.).

i OV-stillinger i islandsk. Dette betyr for det første at språksystemet for både islandsk og norsk reagerer sterkt på faktoren 'informasjonsstatus'. For det andre tyder tallene på at det allerede på et relativt tidlig tidspunkt finnes en syntaktisk forskjell mellom de to språkene. Når man da også vurderer språkhistorien til norsk, så kan den svakere effekten av informasjonsstatus på ordstilling forstås som en indikator og en del av en større syntaktisk utvikling, nemlig en økende fiksering av strikte VO-stillinger i dette språket.<sup>22</sup> Som et resultat av dette er informasjonsstatusen til objekter mindre reflektert i ordstillingen. Resultatene for gammelnorsk i dette prosjektet er påfallende, men det er viktig å huske at KoNoKs-korpuset kun representerer én tekst fra gammelnorsk tid. Med hensyn til dette kan analysen av materialet i KoNoKs-korpuset bare tilby foreløpige resultater.<sup>23</sup> Likevel antyder en slik detaljert analyse av tallene fra ulike syntaktiske mønstre i språket at en divergerende syntaktisk utvikling i gammelislandsk og gammelnorsk skjedde allerede på 1200-tallet og dermed nokså tidlig i språkhistorien. Denne tilnærmingen til materialet med en sterk komparativ ansats bør oppmuntre til flere detaljerte analyser som tar for seg direkte sammenligning av gammelnorsk og gammelislandsk syntaks.

#### 4.2 *Objektsposisjon: Informasjonsstatus og prosodisk vekt*

I studien av objektsposisjonen i gammelnorsk (se Tiemann 2022) som bare er kort skissert her, undersøkes også prosodisk vekt og dens samspill med informasjonsstatus. For gammelislandsk beskriver Thorbjörg Hróarsdóttir (2009) at prosodisk vekt er den avgjørende faktoren for ordning av setningselementer (jf. også Eiríkur Rögnvaldsson 1996: 70), som også kan overstyre informasjonsstrukturelle effekter.<sup>24</sup> Hun viser at fulle

22. Språkhistorisk spiller kاسوبortfall i norsk en rolle i at syntaksen blir mer restriktiv enn i islandsk. Men med et kاسوبystem som fremdeles er aktiv i gammelnorsk blir dette bare en tilleggsfaktor som gjennom tid kan være interessant å se på i samspill med informasjonsstruktur og prosodisk vekt.
23. Siden kun én tekst er representert og annotert her, kan man ikke unngå mulige sjangereffekter. Resultatene må således testes for bredere konklusjoner på grunnlag av analyse av flere gammelnorske tekster fra ulike tidsperioder og ulike sjangere. Det er kjent fra studier av andre eldre språk at tekster kan være nokså ulike med hensyn til ordstilling og informasjonsstruktur.
24. Andre faktorer som ikke er gjort rede for her er rollen av kasus (A-movement; se Chomsky 1981 og Thorbjörg Hróarsdóttir 2000: 115f.). Også Taylor og Pintzuk (2012 a, b) nevner en korrelasjon mellom kasus og ordstilling for gammelengelsk, der objekter i dativ og genitiv oftere står i postverbal posisjon.

(tunge) DP-er oftere står i VO-stilling enn pronomener og lette DP-er, som er foretrukket i en preverbal posisjon.<sup>25</sup> Pronomener viser en tverrspråklig tendens til å stå i en høyere posisjon i den syntaktiske strukturen enn fulle NP-er og DP-er. På bakgrunn av dette forventes det for gammelnorsk at tunge objekter står i VO-stilling og lette objekter i OV-stilling. Det forventes videre at prosodisk vekt fungerer som en blokkeringskonisjon for den informasjonsstrukturelle utløseren, noe som resulterer i at kjente, men tunge objekter blir stående i en VO-stilling.

Prosodisk vekt ble analysert ved hjelp av to separate steg i studien. I en første analyse ble objektene delt inn i tre kategorier: 1) pronomener og demonstrativer (lette objekter), 2) NP-er (nakne substantiver, mellomtunge objekter), og 3) fulle DP-er og PP-er (tunge objekter). I en alternativ analyse ble objektene delt inn i to grupper på basis av stavelsestall: lette objekter med 1–2 stavelser og tunge objekter med 3–6 stavelser. Stavelsestall er målt fra nukleus til nukleus. Objekter som klassifiseres som egen setningsfrase, følger hoved verbet obligatorisk og ble ekskludert fra de to vektanalysene. Tabell 3 og 4 viser resultatene av samspill mellom informasjonsstatus og prosodisk vekt i gammelnorsk (se Tiemann 2022 for gammelnorske eksempler).

	Pronomener og demonstrativer			NP-er			Fulle DP-er og PP-er		
	OV	VO	%OV	OV	VO	%OV	OV	VO	%OV
Kjent	45	18	71,43 %	14	14	50,00 %	18	57	24,00 %
Ny	0	0	0 %	11	28	28,21 %	5	33	13,16 %

Tab. 3: Samspill mellom informasjonsstatus og prosodisk vekt (objekttype)

	Lette objekter (1–2 stavelser)			Tunge objekter (3–6 stavelser)		
	OV	VO	%OV	OV	VO	%OV
Kjent	56	30	65,12 %	21	59	26,25 %
Ny	11	15	42,31 %	5	46	9,80 %

Tab. 4: Samspill mellom informasjonsstatus og prosodisk vekt (stavelsestall)

Resultatene viser at vekt er en signifikant faktor for ordstilling i gammelnorsk. Den samme tendensen kan leses ut av begge de to tabellene,

25. DP betyr ‘determiner phrase’ som har en determinativ som kjerne, ofte med en nominalfrase (NP) som komplement.



nemlig at lette elementer favoriserer en posisjon høyere i den syntaktiske strukturen og dermed oftere står preverbalt. I motsetning til dette står tyngre objekter oftere i VO-stilling. Disse resultatene er også i tråd med en tverrspråklig tendens som viser at lette elementer står foran tunge elementer (jf. Behaghel 1932; se også Hawkins 1994 som argumenterer for at all variasjon i ordstilling er et resultat av syntaktisk vekt). Det er også vanlig at kjente konstituentar i en diskurs er realisert som lette elementer (slik som pronomener), mens nye og fokuserte konstituentar er uttrykt med tunge elementer (slik som substantivfraser). Mer konkret viser resultatene at et kjent objekt som er lett ifølge dets vektanalyse (objekttype/stavelsestill), hovedsakelig står i OV-stilling. Men når et kjent objekt klassifiseres som relativt tungt, står dette hovedsakelig i postverbal stilling. Nye objekter (både tunge og lette) viser derimot en generell tendens for VO-stilling (unntatt pronomener og demonstrativer som vanligvis refererer til kjent informasjon, og derfor ikke forekommer som nye entiteter), se tab. 3. Vekt er altså en signifikant faktor som til en viss grad kan overstyre informasjonsstrukturelle forhold. Dette betyr at selv om OV er foretrukket med kjent informasjon, så forhindrer vektbetingingen en preverbal posisjon hvis objektet er tungt. I tillegg viser tallene at lette objekter i gammelnorsk har en tendens til å være fordelt på VO/OV på basis av informasjonsstatus. Disse følger altså ikke prinsippet om vektfordeling der alle lette objekter skulle stå i en preverbal posisjon. Og her skiller gammelnorsk seg igjen fra gammelisländsk. Thorbjörg Hróarsdóttir (2009: 265f.) skriver nemlig for gammelisländsk at

[L]ight objects with old information (unfocused) are always preverbal and heavy objects with new information (focused) are always postverbal. Light objects with new information are also usually preverbal, while heavy objects with old information are usually postverbal.

Lette objekter i gammelnorsk viser en informasjonsstrukturell effekt og kan stå både pre- og postverbalt, mens tunge objekter oftere dukker opp i postverbal stilling – uansett om de bærer kjent eller ny informasjon. En oppsummering av dette er gitt i tabell 5 og 6 (igjen: tallene er sett i relasjon til VO stilling; asterisken i tabell 5 viser hvor det er forskjell sammenliknet med gammelisländsk).

Objekt	Lett	Tung
Kjent	Preverbal	Postverbal
Ny	*Postverbal	Postverbal

Tab. 5: Oppsummering informasjonsstatus og prosodisk vekt.

%OV	Lett	Tung
Kjent	65,1 %	26,3 %
Ny	42,3 %	9,8 %

Tab. 6: Sannsynlighet for OV-stilling (prosentfordeling).

## 5 Oppsummering: Gammelt materiale og nye korpusanalyser

Denne artikkelen har hatt som mål å vise hvordan det gamle materialet som er arkivert i Språksamlingene, kan brukes i ny forskning. Artikkelen redegjør for omdanningen fra nedskrevet informasjon på papir til XML-koding og videre taggingsprosesser. På grunnlag av den grammatiske merkingen som ble gjort på ordbokssedlene, var det mulig å gjennomføre en omfattende annotasjon av språklige data innen rammene av et PhD-prosjekt. Digitalisering og tilgjengeliggjøring av slikt materiale bidrar til å utvikle nye korpus og til å analysere og dokumentere språket med nye metoder og tilnærminger.

Seddelarkivet representerer slik en nyttig og viktig ressurs der gjenbruk av gammelt arbeid gir en stor gevinst når det gjelder forskningstid og arbeidskraft. PhD-prosjektet om syntaks og informasjonsstruktur i gammelnorsk viser at det er behov for mer detaljerte studier om norsk språk og språklig utvikling, og fremhever også behovet for komparative næranalyser med gammelislandsk. Her må det selvfølgelig trekkes frem at kun én tekst fra den gammelnorske perioden ble analysert i detalj i løpet av dette prosjektet, og at resultatene for denne teksten ikke nødvendigvis representerer den språklige tilstanden for alt materiale fra gammelnorsk tid. Men de skisserte resultatene tyder på noen interessante forskjeller som kun er tilgjengelige gjennom detaljerte korpusanalyser – noe vi følgelig trenger flere av. Denne artikkelen vil altså oppmuntre forskere til videre undersøkelser av gammelnorsk syntaks, separat fra, men sammenlignet med gammelislandsk.

## Referanser

- Bech, Kristin og Kristine, Eide (red.). 2014. *Information Structure and Syntactic Change in Germanic and Romance Languages*. Amsterdam: John Benjamins.
- Behaghel, Otto. 1932. *Deutsche Syntax*. Vol. 4. Heidelberg: Winter.
- Chafe, Wallace. 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics and Point of View. *Subject and Topic*, red. Charles Li, 27–55. New York: Academic Press.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
- Christoffersen, Marit. 1994. Sentences with Initial Adverbials in the Law of Magnus Lagabøter with Particular Emphasis on the Position of the Subject. I: *Language Change and Language Structure. Older Germanic Languages in a Comparative Perspective*, red. Toril Swan, Endre Mørck og Olaf Jansen, 75–90. Berlin/New York: Mouton de Gruyter. [Trends in Linguistics. Studies and Monographs 73].
- . 2002. Har leddsetninga i norrønt ‘forfelt’? Et studium av underordna setninger i Magnus Lagabøters landslov. *Norsk Lingvistisk Tidsskrift* (20), 153–176.
- Donhauser, Karin. 2007. Zur Informationsstrukturellen Annotation Sprachhistorischer Texte. *Sprache und Datenverarbeitung* (31), 39–45.
- Eiríkur Rögnvaldsson. 1994/1995. Breytileg orðaröð í sagnlið. *Íslenskt mál og allmenn málfræði* (16–17), 27–66.
- . 1996. Word Order Variation in the VP in Old Icelandic. *Working Papers in Scandinavian Syntax* (58), 55–86.
- Elenbaas, Marion og Ans van Kemenade. 2014. Verb particles and OV/VO in the history of English. *Studia Linguistica* (68.1), 140–167.
- Faarlund, Jan Terje. 2004. *The Syntax of Old Norse. With a Survey of the Inflectional Morphology and a Complete Bibliography*. Oxford: Oxford University Press.
- Flom, George. 1921–1923. The Language of the Konungs skuggsjá (Speculum regale). *University of Illinois Studies in Language and Literature* (7.3), 1–152; (8.4), 155–323.
- Franco, Irene. 2010. Issues in the syntax of Scandinavian embedded clauses. *Working Papers in Scandinavian Syntax* (86), 137–177.
- Halldor Sigurðsson. 1985. *Diachronic Icelandic Syntax: Some Practical and Theoretical Problems*. Ms., University of Stockholm.

- . 1988. From OV to VO: Evidence from Old Icelandic. *Working Papers in Scandinavian Syntax* (34), 1–41.
- . 1994. *Um frásagnarumröðun og grundvallarorðaröð í forníslensku*. Reykjavík: Bókaverzlun Ársæls Árnasonar.
- Haugen, Jens. 2000. *Old Norse Word Order and Information Structure*. Doktoravhandling. Norwegian University of Science and Technology.
- Haugen, Odd Einar og Åslaug Ommundsen. 2010. *Vår eldste bok: Skrift, miljø og biletbruk i den norske homilieboka*. Oslo: Novus.
- Hawkins, John. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hinterhölzl, Roland og Ans van Kemenade. 2012. The interaction between syntax, information structure and prosody in word order change. *The Oxford Handbook of the History of English*, red. Terttu Nevalainen og Elizabeth Traugott, 803–821. Oxford: Oxford University Press.
- Holmberg, Anders og Christer Platzack. 2008. The Scandinavian languages. I: *The Oxford Handbook of Comparative Syntax*, red. Guglielmo Cinque og Richard S. Kayne, 420–458. Oxford: Oxford University Press.
- Holm-Olsen, Ludvig. 1952. *Håndskriftene av Konungs skuggsjá: En undersøkelse av deres tekstkritiske verdi*. København: Munksgaard. [Bibliotheca Arnamagnæana 13].
- Kruijff, Geert-Jan og Denys Duchier. 2003. Information structure in topological dependency grammar. *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics* (1), red. Ann Copestake og Jan Hajic, 219–226. New York: ACM Press.
- Kruijff-Korbayová, Ivana og Mark Steedman. 2003. Discourse and Information Structure. *Journal of Logic, Language and Information* (12.3), 249–259.
- Lambrecht, Knut. 1994. *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.
- Levinson, Stephen. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- Matić, Dejan. 2015. Information Structure in Linguistics. *International Encyclopedia of the Social & Behavioral Sciences* (12), 2. ed., 95–99.
- Mørck, Endre. 2016. Syntaks. I: *Norsk språkhistorie I: Mønster*, red. Helge Sandøy og Agnete Nesse, 317–445. Oslo: Novus.

- Petrova, Svetlana, Michael Solf, Julia Ritz, Christian Chiarchos og Armir Zeldes. 2009. Building and using a richly annotated interlinear diachronic corpus: The case of Old High German Tatian. *Traitement Automatique des Langues* (50.2), 47–71.
- Pfaff, Alexander. 2019. Reunited after 1000 Years. The Development of Definite Articles in Icelandic. *Nordic Journal of Linguistics* (42.2), 165–207.
- Roberts, Craige. 2012. Information Structure: Afterword to Roberts (1996). I: *Semantics and Pragmatics* 2012 (hefte 5), 1–19.
- Schnall, Jens Eike og Rudolf Simek. 2000. *Speculum regale. Der altnorwegische Königsspiegel (Konungs skuggsjá) in der europäischen Tradition*. Wien: Fassbaender. [Studia Mediaevalia Septentrionalia 5].
- Struik, Tara og Ans van Kemenade. 2018. On the givenness of OV word order: A (re)examination of OV/VO variation in Old English. *English Language & Linguistics* (24.1), 1–22.
- Taylor, Ann og Susan Pintzuk. 2012 a. Rethinking the OV/VO Alternation in Old English: The Effect of Complexity, Grammatical Weight, and Information Status. *The Oxford Handbook of the History of English*, red. Terttu Nevalainen og Elizabeth Traugott, 199–213. Oxford: Oxford University Press.
- . 2012 b. Verb Order, Object Position and Information Status in Old English. *York Papers in Linguistics Series* (2.12a), 29–52.
- Thorbjörg Hróarsdóttir. 1996. The Decline of OV Word Order in the Icelandic VP: A Diachronic Study. *Working Papers in Scandinavian Syntax* (57), 92–141.
- . 2000. *Word Order Change in Icelandic: From OV to VO*. Amsterdam: John Benjamins.
- . 2009. Information Structure and OV order. *Information Structure: Theoretical, Typological and Experimental Perspectives*, red. Malte Zimmermann og Caroline Féry, 258–281. Oxford: Oxford University Press.
- Thorsteinn Indriðason. 1987. *Skýrsla um orðaröð í sagnlið*. Ms., University of Iceland, Reykjavík.
- Tiemann, Juliane. 2018. Nå kan en ny del av det gammelnorske materialet publiseres i digital form. *På Høyden* (jan. 2018). <https://pahoyden.khrono.no/debatt-gamalnorsk-lle/na-kan-en-ny-del-av-det-gammelnorske-materialet-publiseres-i-digital-form/392014>.

- . 2022. The Object Position in Old Norwegian: An Interplay between Syntax, Prosody, and Information Structure. I: *Language Change at the Interface. Intrasentential and intersentential phenomena*, red. Nicholas Catasso, Marco Coniglio og Chiara De Bastiani, 61–94. Amsterdam: John Benjamins.
- . Upublisert manuskript. Syntactic Variation and Information Structure in Old Norwegian: An Investigation of *Konungs skuggsjá* in AM 243 bǫ fol. Upubl. doktoravhandling. University of Bergen.
- Walkden, George. 2014. Object position and Heavy NP Shift in Old Saxon and beyond. *Information structure and syntactic change in Germanic and Romance languages*, 313–340. Amsterdam: John Benjamins.
- Wallenberg, Joel C., Anton Karl Ingason, Einar Freyr Sigurðsson og Eiríkur Rögnvaldsson. 2011. *Icelandic Pared Historical Corpus (IcePaHC)*. Version 0.9. [http://www.linguist.is/icelandic\\_treebank](http://www.linguist.is/icelandic_treebank).

## Abstract

This article describes the life cycle of data based on the physical material in the Language Collections (Språksamlingene) at the University of Bergen. The material was compiled on index cards over Old Norse/Old Norwegian which was later transferred to digital solutions. It is outlined how this material has been reused to build up and supplement a new digital corpus (KoNoKs = Korpus over den norske *Konungs skuggsjá*) designed to answer new research questions. As an example for a new analysis based on this older research material, this article reports on some results of a corpus study on syntactic variation in Old Norwegian.

Juliane Tiemann  
Universitetet i Bergen  
Institutt for lingvistiske, litterære og estetiske studier  
HF-bygget, Sydneplassen 7  
NO-5007 Bergen  
juliane.tiemann@uib.no



# Om *å* som infinitivpartikkel og ei grammatikk-teoretisk fundert leksikografisk formidling

Av Leiv Inge Aa

Artikkelen har to formål: å kartlegge den teoretiske statusen til «infinitivmerket» *å* og å argumentere for ei grammatikk-teoretisk fundert leksikografisk formidling. Etter tusenårsskiftet har *å* vorte grundig behandla både diakront og synkront, og det er ei viss semje om dei diakrone statusendingane: frå subjunksjon i norrønt til verbklitikon i eldre nynorsk tid til subjunksjon igjen frå nyare nynorsk og frametter. Den synkrone subjunksjonsstatusen blir som hjå Faarlund (2015) problematisert her og ein lågare plassert infinitivpartikkel foreslått. Parallelt med den nye *å*-forskinga har det vorte produsert fleire leksikografiske spaltemeter i Noreg enn nokon gong før, og den grammatiske sida av leksikografien har vorte teoretisert. Hovudlinjer frå den grammatisk-leksikografiske litteraturen blir evaluert og med det eit mønster for leksikografisk formidling av grammatiske problem som *å* og andre funksjonsord fører med seg, foreslått. Hovudbodskapen er at ordartikkelen må vere både presis og open nok til å predikere unemnd empiri.

## 1. Innleiing<sup>1</sup>

«Infinitivmerket» *å* er ikkje berre eit infinitivmerke. Sjølv om termen var etablert i skulegrammatikken gjennom mange år (og kanskje framleis er den folk flest brukar), er *å* brukt framfor meir enn verbformer i infinitiv. Derfor kan termen synast empirisk villeiande. I grammatiske innføringsbøker vart *å* omkategorisert til underordna konjunksjon / subjunksjon på 1980–90-talet; sjå t.d. Theil Endresen (1988), Nordgård

1. Eg vil takke *Maal og Minne* sine to anonyme fagfellar, som begge har kome med fruktbare teoretiske, empiriske og metodologiske innspel. Det same gjeld redaktør Hans-Olav Enger, som har bidratt mykje til å betre teksten både fagleg og formmessig.

& Áfarli (1990), Rønhovd (1993), Faarlund (1995). Seinare tilrødde òg Norsk språkråd [Språkrådet] (2005) subjunksjon som skulegrammatikk-term. «Infinitivsmärke» blir framleis brukt som nemning for enkeltordet *å* (jf. Mørck 2019), men sjeldnare som ordklassekategori. I ordbøker er det likevel først i dei siste åra at subjunksjonskategorien er tatt i bruk. I *Norsk Ordbok* (NO), band 12 (2016), vart *å* omtala som infinitivsmärke, og det står framleis i nettutgåva òg.<sup>2</sup> NAOB fører meir informativt opp «subjunksjon; tradisjonelt: infinitivsmärke» medan *Nynorskordboka* (NOB) og *Bokmålsordboka* (BOB) gjer om frå tradisjonell til nyare kategori i det pågåande revisjonsprosjektet (2018–23).

Samtidig med at ordbøkene har streva med å tilpasse seg den nye termtida, har det pågått mykje forvitneleg empirisk og teoretisk diakron og synkron *å*-forskning i Noreg. Frå kring tusenårsskiftet til i dag er det fleire som kastar lys over den diakrone og synkrone statusen til *å*, og sjølv om det er usemje om detaljane, er det ein viss konsensus om at *at/å* har gjennomgått ei statusendring frå subjunksjon i norrønt til å bli grammatikalisert til v-klitikon i eldre nynorsk for så å bli degrammatikalisert til subjunksjon i nyare nynorsk igjen. Men som vi skal sjå, er heller ikkje den synkrone subjunksjonsstatusen uproblematisk og openberr.

Ut av dette spring det to hovudperspektiv som eg vil sjå nærmare på her. For det første vil eg prøve å avklare den synkrone statusen til *å*. Det vil eg gjere gjennom både den føreliggande diakrone og synkrone *å*-forskninga, og gjennom data som Språksamlingane kan bidra med og *har* bidratt med under mitt eige arbeid med det relevante ordbokoppslaget i NO. For det andre vil eg reise meir prinsipielle spørsmål om kva som er ei ordboks oppgåve å formidle om funksjonsord. Det mest interessante med funksjonsorda er kanskje dei grammatiske strukturane orda inngår i, snarare enn funksjonsordet åleine. På den måten kan sjølv ordbokoppsettet, som krinsar om einskilde oppslagsord, synast paradoksalt i desse tilfella. Og når grammatikken er studieobjektet, er det ikkje berre dei frekvente bruksmønsterane som er viktigast å dokumentere – men òg den marginale grammatikken. «Ytterkantane» rissar opp skiljet mellom

2. NO-arbeidet pågjekk over så mange år (dei trykte bøkene kom mellom 1966 og 2016) at både ordtilfang og terminologi endra seg undervegs. Faarlund (2010: 314) meiner at 2014-prosjektet på 2000-talet kunne tatt i bruk dei nye termene frå *Norsk referansegrammatikk* (Faarlund, Lie & Vannebo 1997) medan Aa (2011: 264) argumenterer for at ein burde fullføre papirutgåva med dei gamle ordklassenemningane, men kunne vurdere ei moderat omlegging i nettutgåva.



mogleg og umogleg språk, i dette tilfellet kva som er gangbar norsk og ikkje. Deskriptive ordbøker må definitivt ta ansvar for å beskrive ytterkantane medan normative ordbøker i større grad fremjar prototypisk bruk. Derfor har vi ein potensiell konflikt eller eit paradoks her som framfor alt leksikografar må ta stilling til. Men òg språkfagkunnige ordbokbrukarar bør vere merksame på kva ei ordbok kan, vil og eventuelt bør formidle.

I del 2 vil eg presentere sentrale data som danner grunnlaget for dei ovannemnde problema. Del 3 tar føre seg *å*-forskning frå sirka dei 20 siste åra (3.1), og tilsvarende tidlegare forskning på funksjonsordframstillingar i ordbøker (3.2). I del 4 utdjupar eg to grammatiske ytterkantar ved *å* som ordbøker òg må ta stilling til. I somme tilfelle tar ikkje *å* infinitiv i det heile (og «infinitivsmerke» er såleis utelukka), og i andre tilfelle står *å* bakom setningsadverbiala (og subjunksjonskategorien er problematisk). I del 5 tar eg desse problema vidare i ein leksikografisk kontekst og spør korleis ordbøker skal formidle grammatikk. Eit generelt svar er at dei uansett teoretisk innretning bør etterstreve ein open og predikerande artikkelstruktur. I del 6 rundar eg av.

## 2. Noko av *å*-mangfaldet – sentrale data

*Å* står i ei rad ulike konstruksjonar, og eg vil her skissere noko av mangfaldet, men legge vekt på eit utval av dei dataa som er mest sentrale for diskusjonen. *Å*-mangfaldet kjem tydeleg til uttrykk i dokumentasjonsordbøker, der mange og ulike bruksdøme er samla kompakt. Det vanlegaste for *å* er å innleie infinitivskonstruksjonar, noko som er empirisk udiskutabelt og sjølvstøtt opphavet for termen og den tidlegare ordklassen *infinitivsmerke*. I NO syner dei første døma infinitivskonstruksjonen som objekt (1a), subjekt (1b) og predikativ (1c) i hovudsetninga.

- 1) a. Ingen torde å yppe seg (NO, tyd. A1).
- b. Å trøste enke var eit kjent ord (NO, tyd. A2)
- c. Kvinnfolk e vel ikkje heilt å stola på! (NO, tyd. A4)

Eit teikn på at *å* innleier infinitivskonstruksjonen og ikkje er eit verbklikton, får vi straks infinitivskonstruksjonen er splitta. Døma i (2) er henta frå *å* (A16) i NO; (2a–c) svarar utanom splittinga til den same syntaktiske

bruken som i (1a–c). Oftast blir infinitivskonstruksjonen splitta av ein negasjon, men ikkje alltid, jf. (2b, d). Og oftast er han splitta av eitt ord, men av og til av ein frase, jf. (2d).

- 2) a. Ho lova å ikkje seia noko (NO, tyd. A16a)
- b. Å framleis gi høve til ein tredjedel privat eige står i strid til konsolideringsmodellen (NO, tyd. A16b).
- c. Det beste for sjelefreden er å ikkje ha noko meining (NO, tyd. A16d).
- d. Her må ein ringje på for å i det heile kome seg inn døra (NO, tyd. A16e).

Når NRG tok i bruk subjunksjonskategorien, var det mellom anna fordi kontrollinfiniteivar som i (1) har mykje til felles med finite *at*-setningar. I infiniteivar er det vanleg å rekne med eit usynleg subjekt (ofte markert som PRO) som er tolkbart frå konteksten. I (2a) er PRO kontrollert av subjektet *ho* i oversetninga (jf. NRG: 998). Vidare har infinitivssplitt rådd i norsk lenge og dominerer over ikkje-splitt (meir om dette i 3.1). Gitt at *å* er subjunksjon, skulle ein vente at ordet hadde fleire oppgåver enn å innleie infiniteivar. NRG legg likevel ikkje opp til det, men lanserer i staden ei skriftspråkleg avgrensing som blir dobbelt problematisk dersom ein inkluderer munnleg norsk:

Infinitivskonstruksjonar er konstruksjonar med verbalet i infinitiv og med *å* som innleiarord. Dette er einaste moglege innleiarord, og subjunksjonen *å* har heller ikkje andre funksjonar i språket enn å innleie infinitivskonstruksjonar (NRG: 994).

For det første er *te* 'til' tradisjonelt brukt dialektalt framfor infiniteivar mange stader vestpå (Aasen 1850, 1873, NO *til*, tyd. C11), og som vi ser i (3), kan *å* innleie meir enn infinitivskonstruksjonar. Her følger ei partisippform av verbet, som dessutan har kontrafaktisk innhald. Døme som (3) er veletablerte i norsk; det eldste dømet eg har funne, er (3a) (frå *Syn og Segn* 1911; *Setelarkivet*, SA<sup>3</sup>). Eit enkelt søk i *Nynorskorpuset* gir oss dessutan mange tilslag både frå skjønnlitteratur og sakprosa. (3b) er frå

3. Denne og dei tre *Språknytt*-kjeldene rett nedanfor er søkbare i *Setelarkivet*; sjå SA under «Ordbøker og andre empiriske kjelder» etter den faglege litteraturlista.

avisa *Strilen* (1984).<sup>4</sup>

- 3) a. Hadde dette synet kome til å rådt, då hadde dei europæiske stor-  
samfund fått ein organisk vokster (NO, tyd. B1a)
- b. Det skulle vore hyggeleg å møtt henne att (NO, tyd. B1b)

I NO er desse døma plasserte i ein felles kategori med skildringa «fram-  
for partisipp (el konjunktiv), serl med kontrafaktisk innhald», ei skildring  
som Askedal (2016) kritiserte, men som Wetås & Aa (2017) forsvarte.  
Askedal forfekta (til liks med t.d. Norsk språkråd i *Språknytt* 2/1984: 19;  
SA) eit tradisjonelt syn med ein underforstått og stroken infinitiv. Men  
Eide (2011) syner at kontrafaktiske partisipp kan stå i finitt V<sub>2</sub>-posisjon  
i andre samanhengar, dvs. i C i eit standard CP–TP–VP-system.<sup>5</sup> I (4a)  
er det ikkje mogleg å tenke seg eit stroke hjelpeverb; det går ikkje opp,  
som synt i (4b).

- 4) a. Da <sub>C</sub>fått du en anna opplevels (Oppdal; Ommeren 2010)
- b. Da <sub>C</sub>hadde du fått ... → \*Da \_\_\_\_\_ du fått ...

Eide (2005: 374; 2011: 2) meiner at det kontrafaktiske partisippet er ein  
konjunktiv som berre har form eller ser ut som partisipp; V<sub>2</sub>-distribusjo-  
nen gjer det faktisk uråd å kategorisere *fått* i (4a) som partisipp. Truleg er  
vi her i skjeringspunktet mellom det normative og det deskriptive i den  
norske grammatikken. Sandøy (1991) koplar liknande døme eksplisitt til  
føringar i den normative skulegrammatikken. Vi får innprenta i skulen  
at *å* må ta infinitiv, og det påverkar òg informantar i relevante utspørjingar.

4. Liknande finn ein òg i færøysk. (i) er henta frå nettavisa *Dimmalætting*  
(<https://dimma.fo>, 08.10.21). Takk til eine fagfellen for denne merknaden.  
(i) Vit áttu at havt vunnid 10–1 í Landskrona.
5. For dei som er ukjende med dette systemet, finst det ei enkel og fin innføring i Áfarli  
(2019). Der blir modellen sine føremoner òg jamførde med Diderichsen sitt tradis-  
jonelle feltskjema. Grovt (men litt unøyaktig) kan vi seie at C-frasen (CP; C = eng.  
*complementizer* 'subjunksjon') svarar til setningas forfelt, T(empus)-frasen til midt-  
feltet (der verbet ev. blir tilordna tempus) og V(erb)-frasen til predikatet (hovud verbet  
med objekt og predikatsadverbial). CP–TP–VP modellerer setninga hierarkisk (ulikt  
feltskjemaet), og kvar av frasane har ein kjerne. Kjernen i CP er C, og dette er posis-  
jonen til subjunksjonen i undersetningar og V<sub>2</sub>-posisjonen i hovudsetningar (dit det  
finitte verbet flyttar). At subjunksjonen og det finitte verbet «konkurrerer» om den  
same posisjonen i setninga, kjem ikkje fram av feltsjemaanalysen.

Sandøy meiner at skuleidealet gjer «at vi til dels mistar språkkjensla for morsmålet på dette punktet» (s. 259). Som Wetås & Aa (2017) syner, går den normative koplinga mellom *å* og infinitiv tilbake til Aasen (1864) (og kanskje lenger òg), der han m.a. påpeiker at ein bruk som i (3) finst i dialektane, «men den er noget uformelig og burde neppe innføres i skrift, hvor man hellere maa bruge det fuldkomne Infinitiv ...» (§ 324, anm.). Deskriptivt er det ikkje råd å snakke om eit infinitivsmärke her, og skal ein likevel forfekte det, tilslører ein eit viktig register vi har som norske morsmålsbrukarar, i skrift. Det gjer etter mitt skjøn skriftspråket fattigare. Særleg bør ein ha dette i mente for nynorsken sin del, dersom språket i det heile skal ha kontakt med folkemålet og si eiga skrifthistorie. Nynorskskrivarar inkluderer sjølvsagt talarar med ulike dialektar og såleis med ulike internaliserte grammatikkar. Denne skrivaren er vestlending, og for eigen del er døma i (3) heilt gjengs medan (4a) er umogleg. Vi må ta høgde for at nynorsk lar seg realisere med ulike grammatiske system. Om statusen til *å* er ulik i dei ulike systema, kjem vi tilbake til i del 4.

Det finst òg data som ulikt (2) og (3) gjer subjunksjonskategorien problematisk. Dei mest konvensjonelle døma som talar imot, er «ikkje-splitta» infinitivskonstruksjonar (5a), som jamvel har vore det normativt tilrådde i skrift i mange år. Det må sjåast i samanheng med at dette er i pakt med det danske mønsteret. Norsk språkråd rådde ifrå splitting i *Språknytt* (3/1981: 11; SA) medan dei differensierte mellom bokmål (ikkje-splitt) og nynorsk (splitt) ti år seinare (*Språknytt* 2/1991: 19; SA). Her syntest ein altså å meine indirekte at ikkje-splitt er danskavleidd medan splitt er vanlegare i norsk folkemål. I alle høve er (5a) fullt mogleg norsk, og då er det problematisk å kategorisere *å* som subjunksjon (C) her. Grunnen til det er at det er umogleg å legge til eit setningsadverbial, *aldri*, på tilsvarande plass, dvs. fremst, i den finitte *at*-setninga i (5b) (setningsadverbialalet kan med andre ord ikkje adjungerast til CP). Begge døma er henta frå Faarlund (2015).

- 5) a. Eg kan ikkje lova [aldri å gjera det igjen].  
 b. \*Eg kan ikkje lova [aldri at eg skal gjera det igjen].

Åfarli & Eide (2003) meiner at vi med dette har to ulike *å* i norsk, noko vi kjem tilbake til i 3.1. Faarlund (2015) bidrar med endå fleire døme som problematiserer subjunksjonskategorien. I (6) har vi dobbelførekomstar av *å*, som nok må reknast som uformelt, men som i alle fall er attesterte.

Døma er igjen Faarlund sine.<sup>6</sup>

- 6) a. Ønsker du å garantert å bli rik, ...
- b. Da er det å bare å glede seg ...
- c. Arbeidsplassen tjener mer på å faktisk å ha folk på jobb.

Til slutt vil vi nemne tilfelle der *å* ikkje tar eit verb i det heile. Dette òg må reknast som uformelt, men er slett ikkje uvanleg munnleg. (7) er kanskje likevel mindre aktuelt for skriftleg norsk.<sup>7</sup> Døma er henta frå Aa (2014: 137); begge er frå SA.

- 7) a. For itte å nevne hvis en skulle ha vøri nødt til å opp på potta om natta.
- b. Ho kjælttest ('var redd for') å ut etter det var mørkt.

Av døma vi skal omtale vidare, har vi då frekvente og konvensjonelle (1), der ein infinitiv følger etter *å*, og likeins frekvente (2), der eit setningsadverbial følger etter *å*. Døma i (3) er mindre frekvente, men viktige når vi skal avklare statusen til *å*, som her ikkje tar infinitiv i det heile. I (5) og (6) har vi høvesvis litt og veldig mykje mindre frekvente døme enn i (2), men til liks med døma i (3) må vi ta omsyn til desse òg. Her kan ein spørje seg i kor stor grad slike ytterkantdøme skal inkluderast i ein teoretisk diskusjon. Det kjem nok m.a. an på kva ein forstår med *språk*. I generativ grammatikk er ein primært interessert i å forstå det internaliserte språket (I-språket) hjå den einskilde språkbrukaren, og frekvens blir då mindre viktig. Korpusdata svarar til fleire språkbrukarars «E-språk» og vil ikkje åleine avsløre dei grammatiske moglegheitene hjå den

6. Faarlund har hausta sine data frå internett (Google), noko fleire har diskutert både kvalitative og kvantitative problem med. I dette tilfellet er dei kvalitative problema mest påtrengande. Schütze (2010: 210) understrekar at ein ikkje alltid veit kven som står bak einskilde nettførekostar, ein veit då heller ikkje kor representativ ein førekomst er, dvs. om fleire vil godta den same setninga. I prinsippet gjeld det siste for (talespråks)korpustreff òg (Aa 2015b: 124).
7. Vi har òg variantar der ein *å* + preposisjon tilsynelatande er koordinert med ein vanleg verbfrase, jf. (i) (frå <https://nasjonalparken.no>; takk til eine fagfellen for dømet).
  - (i) Neste dag stekte sola på teltet tidlig så det bare var å opp og fyre på vatn til havregryna.

Her kan ein kanskje tydelegare enn i (7) sjå føre seg *å opp* som ein «verblaus verbfrase», men ein analyse utan usynleg verb kan fint forsvarast.

einskilde (jf. Fanselow, Féry, Vogel & Schesewsky 2008: 4). I leksikografien er ein tradisjonelt mindre opptatt av språket som ein individuell og internalisert storleik. Eg legg likevel vekt på å inkludere «mentalt funderte ytterkantar» når eg skal prøve å konkludere med den *teoretiske* statusen til *å* i del 4. Leksikologisk forskning verken kan eller bør vere teoretisk eller metodologisk einsarta. Tvert imot må ein fundere beskrivinga av grammatiske ord grammatikkteoretisk dersom ein skal ta på alvor å beskrive denne delen av ordtilfanget òg.

### 3. Tidlegare blick på *å* spesielt og funksjonsord i ordbøker generelt

I denne bolken vil eg diskutere nokre tidlegare arbeid om *å* spesielt (3.1) og tidlegare funksjonsorddiskusjonar i skjeringspunktet mellom grammatikk og leksikografi (3.2). Dette samsvarar òg med rekkefølga på diskusjonsbolkane etterpå, der det *å*-spesifikke blir diskutert i del 4 og overordna grammatiske ordbokimplikasjonar i del 5. Målet er såleis todelte, men må sjåast i samanheng: å få avklara grammatiske eigenskapar ved *å* og å diskutere korleis dette og annan grammatikk kan formidlast i leksikografien.

3.1 *Grammatisk kategorisering og forskning på diakron og synkron bruk av å*  
Som nemnt i innleiinga (del 1) vart *å* omkategorisert til subjunksjon i løpet av 1980–90-åra i grammatisk innføringslitteratur. Faarlund si korte morfologiinnføring kan illustrere overgangen. I førsteutgåva (1988: 38) er *å* berre omtala som infinitivmerke medan i andreutgåva (1995) er ordet plassert blant subjunksjonane (s. 32), slik Rønhovd (1993: 89) hadde gjort to år i førevegen. Men ein fann òg nemninga «(underordnande) konjunksjon» alt hjå Theil Endresen (1988: 133)<sup>8</sup> og Nordgård & Åfarli (1990: 117). Som det går fram av forordet i Faarlund (1995), var systemet der tilpassa NRG (1997), som var under arbeid samtidig. Og NRG vart sjølvstøtt ein referanse alle må forhalde seg til på ein eller annan måte. Det gjorde òg Norsk språkråd [Språkrådet] (2005) når dei utarbeidde termar til bruk i

8. Theil Endresen (1988) er morfologikapitlet i 3. utgåva av *Språkvitenskap. En elementær innføring* (Gram Simonsen, Theil Endresen & Hovdhaugen 1988). I dei to første utgåvene av same boka (Hovdhaugen 1978 og Bjorvand, Hovdhaugen & Gram Simonsen 1982) er infinitivmerket verken nemnt som eigen ordklasse eller blant dei underordna konjunksjonane.

skuleverket. Der er *å* likeins subjunksjon. Vi ser òg at subjunksjonskategorien var etablert i fleire innføringsbøker rundt tusenårsskiftet, som i Åfarli (1997: 150) og hjå dei ulike forfattarane i Bye, Trosterud og Vangnes (2003). Til saman er det derfor rimeleg å seie at *å* «offisielt» har ein synkron status som subjunksjon i norsk, og det er utgangspunktet vidare.

I det følgjande vil eg gjere greie for hovudtrekk i den diakrone utviklinga til *å*, før eg rundar av delkapittelet med Faarlund (2015, 2019) sine synkrone perspektiv. Kort fatta synest det å vere ein konsensus mellom Faarlund (2003, 2005), Myrvoll (2005) og Mørck (2019) om at *at/å* har gjennomgått ei statusendring frå subjunksjon i norrønt til å bli grammatikalisert til *v*-klitikon i eldre nynorsk, for så å bli degrammatikalisert til subjunksjon i nyare nynorsk igjen. Tidsgrensene er ikkje openberre, men Myrvoll og Mørck trekker den siste degrammatikaliseringsgrensa ein god del lenger tilbake (til 1700-talet) enn Faarlund opphavleg gjorde.

Holmberg & Platzack (1995) observerte ein konsekvent skilnad på ordstillinga i undersetningar mellom islandsk og eldre skandinavisk på eine sida og moderne skandinavisk på hi. I islandsk og eldre skandinavisk (med svensk som døme hjå dei) står det finitte verbalet framfor setningsadverbial, medan det står bakom i dei moderne fastlandsspråka. (8) illustrerer (frå Holmberg & Platzack 1995: 75, 77).

- 8) a. **islandsk:** að Jón keypti ekki bókinu.  
 b. **svensk:** att Ulf inte köpte boken.  
 c. **gamalsvensk:** at Gudz ordh kan ey vare j honom

Holmberg & Platzack analyserer forskjellen som at dei moderne fastlandsspråka manglar ei verbflytting forbi setningsadverbialet, ei flytting som islandsk framleis har etter gamalt.<sup>9</sup> Poenget til Faarlund (2003) i

9. Dei ser dette i samanheng med at samsvarstrekket (AGR) må sjekkast lenger oppe i setninga. At det er ein samanheng mellom verbmorfologi og adverbposisjonar, stemmer ikkje for elvdalsk (Garbacz 2010) eller for mange språk globalt (Newmeyer 2005). Som såkalla *makroparameter* er derfor AGR-hypotesen falsifisert. Det kan òg diskutertast om han i det heile fungerer som ein deskriptiv regel for dei nordiske språka (minus elvdalsk), ettersom ein i norsk òg kan ha verbflytting forbi setningsadverbialet i *at*-setningar t.d. ved indirekte tale (i). Med andre subjunksjonar er dette meir problematisk (ii).
- (i) Han sa at Jon (hadde) ikkje (hadde) kjøpt boka.  
 (ii) Han skulle feira dersom Jon (\*hadde) ikkje (hadde) kjøpt boka.

Dermed kan *at*-setningar tilsynelatande ha ein dobbel CP-struktur (til subjunksjon

denne sammenhengen er at skiftet òg viser seg i infinitivar:

- 9) a. **norrønt**: *at* + V + ADV  
 b. **eldre nynorsk**: ADV + *å* + V

Norrønt *at* blir analysert som ein subjunksjon (generert i C), men infinitivkonstruksjonen blir ikkje splitta fordi verbet flyttar ut av VP. Sidan adverbet/negasjonen markerer kanten av VP (Adger 2003),<sup>10</sup> er det klart at verbet blir verande, og *å* må likeins stå inni verbfrasen, klitisert til verbet.

Parallelt med den syntaktiske statusendringa har det skjedd ei fonologisk endring ved at rotvokalen har vorte lengd, runda og korta, og at konsonanten til slutt har falle bort (og vokalen har vorte lengd på nytt). Myrvoll (2005: 276) illustrerer prosessen som i (10).

- 10) /at/ > /a:t/ > /o:t/ > /ot/ > /o:/

Faarlund (2003) ser den fonologiske og syntaktiske endringa i samanheng, og han meiner at *at/aat* må ha vore eit sjølvstendig ord så lenge det var tyngd. Myrvoll viser til problem ved analysen og ser prosessane uavhengig av einannan; det siste steget i (10) har samanheng med eit allment *t*-bortfall. Her trekker han òg fram eit døme frå Teleman (2004) si melding av same Faarlund-boka, som syner at ordet har vore brukt som «infinitivmerke» med forma *å* i svensk òg, men utan at det nokon gong har vore grammatikalisert til eit klitikon. Svenske infinitivkonstruksjonar har stadig vore splitta, og «infinitivmerket» er såleis truleg best analysert som ein subjunksjon uavhengig av fonologisk form. Faarlund føreset dessutan at (9b) representerer «grunnrekkefølga», medan *å* + ADV + V har kome først på 1900-talet. Her opponerer både Myrvoll (2005: 277–278) og Mørck (2019: 109–110) med eintydige data som viser at infinitivkonstruksjonar har vore splitta i norsk mykje lenger. Mørck syner at den moderne varianten (med splitting) i alle fall har levd attmed

og finitt verb) som andre undersetningar ikkje kan ha.

10. Faarlund (2015: 4–5) syner at setningsadverb òg må kunne adjungerast til TP; sjå del 4. Same Faarlund (s. 5–6) syner likevel at norrønt *at* må forståast eintydig som subjunksjon.



den eldre nynorske varianten (utan splitting) sidan 1700-talet.<sup>11</sup> Gjennom dei siste hundreåra har vi då tilsynelatande hatt to variantar av *å*: éin som verbklitikon (11a) og éin som subjunksjon (11b). Som ein ser her, markerer negasjonen kanten av VP og er såleis venstreadjungert til den i begge tilfella, i tråd med det som vart sagt ovanfor. (11a) svarar elles til (5) og (11b) til (2).<sup>12</sup>

- 11) a. foreldra har valt [<sub>CP</sub> [<sub>TP</sub> PRO [<sub>VP</sub> ikkje [<sub>VP</sub> [<sub>V</sub> [<sub>klit</sub> å] [<sub>V</sub> døypa]]] barnet]]]] (NPK 2008; *Nynorskorpuset*)  
 b. ho lova [<sub>CP</sub> [<sub>C</sub> å [<sub>TP</sub> PRO [<sub>VP</sub> ikkje [<sub>VP</sub> [<sub>V</sub> seia] noko]]]]] (NO, tyd. A16a)

Det er to innvendingar mot at *at/å* faktisk har skifta status, slik vi har gått ut frå ovanfor. Den eine er teoretisk, det andre empirisk. For det første kjem det an på korleis ein avgrensar kategorien subjunksjon. Det kan tenkast at det nettopp er subjunksjonen som er klitisert i eldre nynorsk (jf. (9b), dvs. at den fonologisk reduserte realiseringa ikkje endrar på kategorien utover det. Når Myrvoll (2005) og Mørck (2019) syner at splitting går lenger tilbake enn Faarlund (2003) først hevda, får vi ein heller kort mellomperiode der *å* er påstått å ha ein annan grammatisk status (som ikkje-subjunksjon). Og sidan kjeldematerialet i denne mellomperioden attpåtil er tynt, kan mangelen på ordstillinga i (9a) potensielt forklarast nettopp gjennom det mangelfulle tilfanget. Samla kan det altså reisast tvil om at *at/å* har skifta status «att og fram».<sup>13</sup>

Likevel sit vi att med to *å*-posisjonar i moderne norsk, slik at statusen til ordet ikkje er eintydig. Også Åfarli & Eide (2003) observerer dette, og dei forklarar dette utan å ty til ein klitiseringsanalyse. Heller meiner dei at *å* skal genererast i T i (11a). Det betyr like fullt at setningsadverbialet må adjungerast til TP som hovudregel, noko dei òg gjer konsekvent

11. I SA er det òg dokumentert splitting før 1900. I Aa (2014: 136) syner eg døme frå framstående landsmålsforfattarar som Arne Garborg og Aa. O. Vinje og hevdar at splittinga dermed må ha vore godt etablert i folkemålet i samtida. Garborg og Vinje bryt i alle fall med dansk syntaks.
12. Mørck (2019) brukar notasjonen I(inflexion)P i staden for T(empus)P, men eg held meg til TP her i representasjonen av moderne norsk (som hjå Åfarli 2019 og Åfarli & Eide 2003), der nettopp tempus er det relevante bøyingsstrekket å sjekke for verbet. Det blir òg eit poeng for kategoriseringa av *å* lenger nede.
13. Takk til eine fagfellen for gode merknader her.

i systemet sitt.<sup>14</sup> Det er eit liknande utgangspunkt Faarlund (2015) har når han skal forklare dataa med «dobbelte infinitivmerke», jf. (6) ovanfor; (6c) er gjentatt her som (12).

12) Arbeidsplassen tjener mer på å faktisk å ha folk på jobb.

Som i ikkje-splittkonstruksjonane hjå Åfarli & Eide (2003) kan ein her sjå føre seg at *å* først blir spleisa i T, men sidan blir kopiert til C, utan at den nedste kopien blir sletta. Dette kan såleis generaliserast til ein generell flyttings-/kopieringsregel, og det er det Faarlund (2015) gjer. Då vil *å* flytte til C i splittkonstruksjonar, der skilnaden frå (12) er at den nedste kopien blir sletta. Dette betyr at Faarlund ikkje stør den «offisielle» norske kategoriseringa av *å* som subjunksjon, men «infinitivmerke» er òg problematisk gitt døma i (3). Historisk sett er det kanskje ikkje urimeleg at vi har med to ulike *å* å gjere, men eg kjem til å argumentere for at dei berre er to ulike variantar av det same *å*-et (begge er genererte i T, men berre den eine flyttar til C).

### 3.2 Studiar av funksjonsord i ordbøker

Samtidig med at den nyare *å*-forskinga har pågått, har det vore reist både spesifikke og generelle problemstillingar om semantikken til og framstillinga av ulike funksjonsord i ordbøker. Eit tidleg arbeid i norsk samanheng som utforskar landskapet mellom lingvistisk teori og leksikografisk praksis, er Runde & Kristoffersen (2006) sin studie av *til*, som er gjord i ein kognitiv grammatisk tradisjon (jf. Langacker 1987). Eit overordna metodisk spørsmål dei stiller, er om det leksikografiske arbeidet med det som i stor grad er grammatiske problemstillingar, skal vere reint materialstyrt («nedanfrå og opp») eller teori- og hypotesestyrt («ovanfrå og ned»). Dei landar på den siste løysinga, noko eg kjem tilbake til og stør opp om i del 5. I artikkelen utforskar dei *til*-materialet spesifikt (med utgangspunkt i NOB) og preposisjonssemantikk meir generelt. Når dei fremjar ein leksikografisk struktur for preposisjonen, er det med utgangspunkt i kognitiv språkteori, der røyntslene våre med den fy-

14. Det må dei gjere så lenge dei reknar med at verbet prinsipielt flyttar til T for å sjekke tempustrekket – òg i undersetningar. Då blir skilnaden på hovud- og undersetningar at verbet flyttar – forbi setningsadverbialet – til C berre i hovudsetningar. Som nemnt lenger oppe er C «blokkert» av subjunksjonen i undersetningar, slik at verbet ikkje kjem lenger enn til T der (med unntak av i nokre *at*-setningar).

siske verda rundt oss er avgjerande for korleis vi tolkar og strukturerer røyndommen. Og sidan språkleg omgrepsdanning og kategorisering speglar den meir generelle, ikkje-språklege kategoriseringa i denne teorien, blir fysiske eller lokative språkuttrykk rekna som grunnleggande, og andre uttrykk avleidde av desse (Runde & Kristoffersen 2006: 15). Dette tankesettet har prega struktureringa av preposisjonsartiklane i NO i nyare tid, men ikkje i dei første banda.<sup>15</sup>

I Halmøy & Berg-Olsen (2012) blir det synet òg fremja at *som* har ei grunntyding, men ei tyding som er meir abstrakt enn den ovannemnde analysen av *til*. Dei opponerer mot at *som* er eit polysemt ord og fremjar den semantiske grunntydinga predikasjonspartikkel (jf. òg Bowers 1993; Åfarli & Eide 2001), som gjeld for all bruk av ordet. Kort sagt er *som* eit ikkje-verbalt kopula som tilskriv ein entitet (subjektet) ein eigenskap (predikatet), og dette er ei konstant og kontekstuavhengig tyding. Gjennom artikkelen analyserer dei det empiriske mangfaldet i dette lyset og syner med det at det som tradisjonelt er tolka som tydingsmangfald, i staden er kontekstuelt funderte *tolkingar* av den same tydinga. Bouchard (1995: 16–17) definerer dette som G(rammatisk)-semantikk (den invariante tydinga) og S(ituasjons)-semantikk (tolkingane). Tilfelle av polysemi har vi når det S-semantiske mangfaldet kan sporast tilbake på den same G-semantikken, og dette blir for han den mest prinsipielle måten å skilje polysemi frå homonymi på. Bouchard teiknar med dette ein grenseopp- gang mellom semantikk som er i eit ein-til-ein-forhold med den syntaktiske strukturen, og semantikk som ikkje er språkspesifikk, men relaterer til kunnskap om verda. Det siste vil somme beintfram kalle pragmatikk, og fleire vil avvise ei så klar grense mellom dei to nivåa (t.d. Jackendoff 1983). Ein hovudmotivasjon for Bouchard er å kartlegge den delen av semantikken som korresponderer med form. Å avvise eit skilje mellom den grammatisk relevante og den situasjonelle semantikken meiner han gjer forholdet mellom form og innhald vilkårleg, noko som legg opp til ein mykje mindre rigid argumentdistribusjon enn det som er tilfellet (Bouchard 1995: 22).

Dersom vi følger hypotesen om ein einsarta G-semantikk, får den gamle samanlikningskonjunksjonen (13a; *som* tyd. A i NO) spora den

15. Preposisjonen *av* (III av) i NO vart t.d. trykt i band 1 (1966), altså lenge før Langacker (1987). Men også *for* (IX for; trykt første gongen i band 3, hefte 1, 1989) har ein heilt annan struktur enn Runde & Kristoffersen forfektar.

same eigenskapstilordnande tydinga som det vi kan kalle relativsubjunksjonen (13b; *som* tyd. B i NO).

- 13) a. Dei fann han (...) som død (NO, tyd. A1a UT1)  
 b. den stunda som eg var der (NO, tyd. B1)

I (13a) får småsetningssubjektet *han* tilordna eigenskapen *død* direkte gjennom det Halmøy & Berg-Olsen (2012) kallar ein fraseekstern predikasjon. I relativsetninga i (13b) «mettar» *som* relativsetninga, som er kjenneteikna av å ha ein tom posisjon (jf. NRG: 1047, 1054), med korrelatet *den stunda*. Denne predikasjonen omtalar Halmøy & Berg-Olsen som fraseintern. Både Runde & Kristoffersen (2006) og Halmøy & Berg-Olsen (2012) er viktige og nytenkande arbeid som har tatt den grammatiske leksikografien eit langt steg vidare i Noreg. Viktig var òg meldinga til Faarlund (2010) av NO (band 8), der det leksikografiske arbeidet vart vurdert kritisk frå grammatisk hald. Eg prøvde sjølv å følge opp denne meldinga ved å diskutere preposisjonskategorisering i ordbøker (Aa 2011), som var noko av det Faarlund kritiserte i NO. Vidare utover 2010-talet vart *med*, *på* og dels *i* diskutert meir prinsipielt i Aa (2013, 2015a), *mot* i Aa (2017a) og *opp* i Aa (2017b). I tillegg vart *å*-problemstillingar diskuterte heilt kort i Aa (2014), Askedal (2016) og Wetås & Aa (2017).

Preposisjonsoppslag som *med*, *mot*, *mellom*, *opp* og *på* i NO, band 7 (2008) og 8 (2009), har ein grunnstruktur som minner om den Runde & Kristoffersen (2006) forfektar. Dei ovannemnde artiklane av Aa skil seg frå dei relevante preposisjonsoppslaga i NO ved at det blir argumentert for eit system som minner meir om Halmøy & Berg-Olsen (2012) sitt, dvs. at preposisjonane har ei abstrakt grunntyding som speglar all bruken. Til dømes blir det hevda at *med* sin G-semantikk er samanstilling, *på* sin er kontakt og *i* konteinar (Aa 2013; jf. Bouchard 1995: 13). Til skilnad frå Runde & Kristoffersen sin analyse blir G-semantikken hevda å vere meir abstrakt enn lokativ. Det vil seie at den fysiske konseptualiseringa er nettopp det første ikkje-språklege laget vi legg på, men G-semantikken er konstant og uavhengig av kontekstuelle vilkår (S-semantikk). Dette er vidare utprøvd i Aa (2018), der *med* blir diskutert som preposisjon og partikkel i G-semantisk perspektiv. Det overordna poenget i denne artikkelen er at all bruk av preposisjonen og partikkelen (dvs. all S-semantikk) er gjennomsyra av den same G-semantikken: samanstilling. Dette arbeidet

er ikkje direkte ordbokrelatert, men kan likevel sjåast på som ei vidare teoretisk og empirisk utprøving av perspektiva som er kortare greidd ut om i Aa (2013: 155–157).

Samla kan ein konkludere med at den leksikografiske praksisen og den leksikografisk-grammatiske litteraturen dei siste 15 åra har gått i to retningar: ei som er funksjonelt orientert etter Langacker (1987), og ei som spring ut av ein formell teori, særleg etter Bouchard (1995, 2002). Eg vil kome tilbake til det eg meiner er den leksikografiske vinsten av dette, i del 5.

#### 4. Å og grammatiske ytterkantar

I møte med generativ litteratur kan ein somtid stusse over dei empiriske døma som dannar utgangspunktet for eit teoretisk resonnement. Eit slikt døme finn vi innleiingsvis i verbpartikkelklassikarane Kayne (1985) og Dikken (1995), jf. (14).

14) They made John out a liar.

Så langt eg veit, er dette eit døme ikkje alle engelske morsmålsbrukarar aksepterer (jf. Larsen 2014: 20, fn. 7; Aa 2020: 71, fn. 8). Men det er ikkje tilfeldig at diskusjonane hjå Kayne og Dikken spring ut av marginale konstruksjonar og er baserte på nokre få døme som tilsynelatande fell akkurat innanfor vs. utanfor akseptabilitetsgrensa. Metodisk er dette eit bevisst grep for å utforske kva den mentale grammatikken rommar. Og derfor er altså òg dei marginale ytterkantane, døme ein nesten aldri ser i bruk, teoretisk interessante og avgjerande. Ein slik inngang kan verke paradoksar for den meir empirisk funderte leksikografiske tradisjonen, der bruksfrekvens spelar ei rolle for kategoriane/tydingsbolkane som blir laga, og prototypiske døme illustrerer desse kategoriane. Grensene for mogleg språk kan synast mindre relevant. Likevel er leksikografien mangromma; *ord* vil ikkje berre seie ‘innhaldsord’. Ein skal dekke funksjonsorda òg, og då er strukturane som desse orda inngår i, det mest vesentlege å formidle. I leksikografien kallar ein dei nummererte og bokstaverte oppdelingane i ordartikkelen tydingsbolkar, men det kan synast misvisande særleg for funksjonsorda. Gitt at tydinga/G-semantikken er konstant, kan vi heller kalle dei tolkings- eller bruksvariantar (som hjå

Halmøy & Berg-Olsen 2012). Og blant desse bruksvariantane er altså dei marginale variantane prinsipielt like viktige; døma i (3), (5) og (6) er like viktige som døma i (1) og (2). Og då viser det seg nettopp at einskapleg status for ordet i alle samanhengar *ikkje* treng å vere så beintfram enkel som dagens offisielle subjunksjonskategori kan tyde på. Med andre ord: *Er verkeleg å ein subjunksjon?*

I den vidare diskusjonen skal vi særleg inkludere to typar data: 1) Døme der *å* ikkje tar infinitiv (og såleis truleg må forståast som VP-eks-tern, kanskje ein subjunksjon), og 2) døme der nærleiken mellom *å* og infinitiven er så påtrengande at det er vanskelegare å forsvare subjunksjonskategorien. Umiddelbart synest altså *å* så grammatisk heterogent at det ikkje er gitt at det soknar til éin ordklasse. Dersom ordet er rimeleg å kategorisere som polysemisk «på tvers av ordklassar» (Kinn 2007: 159), har vi med tilfelle av heterosemi å gjere (jf. Lichtenberk 1991). Tilsvarende er *som* brukt i grammatisk så ulike konstruksjonar at det passar i forskjellige av dei etablerte ordklassekategoriane. Den oppgitte ordklassen er derfor ikkje ei grammatisk og G-semantisk sameinande nemning; det er derimot *predikasjonspartikkel*. Kva uttrykker så *å*, gitt at det er det same ordet i dei ulike konstruksjonane?

Vi repeterer det kontrafaktiske dømet (3b) som (15).

15) Det skulle vore hyggeleg å møtt henne att (NO, tyd. B1b)

Slike kontrafaktivar med ei partisipial verbform representerer som nemnt ein veletablert konstruksjon i norsk. Det er openbert vanlegare med splitta enn ikkje-splitta infinitivkonstruksjonar, og splittinga er vanlegare i dei partisipiale variantane òg. I splitta (16a) kan vi gå ut frå eit «høgt» *å*, dvs. i C. Kombinasjonen *ikkje + å* kontrafaktivt partisipp er vanskelegare å finne autentiske døme på. I (16b) har vi inkludert eit døme der partisippet ikkje er kontrafaktivt, og frå eit normativt hald er det endå meir nærliggande å gå ut frå ein utelaten infinitiv her.

- 16) a. det [ville] vore unaturleg **å ikkje trykt** ein faksimile for å illustrera kva saka gjaldt (*Dag og Tid* 2006; *Nynorskorpuset*).  
 b. Nissedal kommune har ved fleire høve fått kommentarar frå Fylkesmannen for **ikkje å hatt** vaktberedskap i eit så stort hytteområde (*Vest-Telemark blad* 2006; *Nynorskorpuset*).

Sjølv om det er vanskeleg å finne døme på typen (16b), er konstruksjonen fullt mogleg innanfor norsk grammatikk. I skrift er døma kanskje få sidan dei som er normativt opptatt av å ikkje splitte i første omgang, vil vere tilsvarende opptatt av at bakom *å* skal det følge ei infinitivsform. I tale er splitting mykje vanlegare enn ikkje-splitting, og (16b) blir derfor «dobbelt marginal». I tillegg er talespråkkorpusa sjølvstøtt mykje mindre enn skriftspråkkorpusa, slik at sjansane for finne munnlege kontrafaktiske ikkje-splittdøme samla er små. I praksis får vi derfor eit *å* langt framme (eller høgt oppe; i C) i setninga i kombinasjon med partisipp. Men eit *å* lenger bak i setninga (eller lågare; i T) + partisipp er stadig mogleg norsk grammatikk.

(4a) = (17) syner at kontrafaktiven må forståast som finitt i somme samanhengar, sjølv om han ikkje er det i kombinasjon med *å*.

17) Da <sub>C</sub>fått du en anna opplevels (Ommeren 2010).

Lik infinitivssetningar må vi rekne *å*-setninga i (15) som infinitt, og det er rimeleg å gå ut frå at det nettopp er eit generelt infinitt trekk *å* er assosiert med. Åfarli & Eide (2003) og Faarlund (2015, 2019) har alt synt at *å* ikkje kan forståast eintydig som ein subjunksjon, noko som altså òg er tilfellet når verbforma ikkje står i infinitiv. Om ein reknar med at setningsadverba kan adjungerast både til TP og VP, kan ein gjere greie for både splitt og ikkje-splittdøme med *å* generert i T (Faarlund 2015: 4). Det generelle infinitte trekket er såleis T-assosiert. Likevel finn Faarlund historiske argument for at *å* må kunne kopierast til C. Norrønt *at* var eintydig subjunksjon (Faarlund 2015: 5–6), og likeins ser ein med døma i (6) ovanfor at ikkje berre setningsadverbial, men *å* har to moglege posisjonar, som i (18) = (6c).

18) [<sub>CP</sub> [<sub>C</sub> *å* [<sub>TP</sub> faktisk [<sub>TP</sub> PRO [<sub>T<-FIN></sub> <*å*> [<sub>VP</sub> ... ha ...]]]]]]

I denne analysen vil då den nedste kopien normalt strykast; skilnaden på splitta og «doble» infinitivar er nettopp stryking vs. bevaring av kopien i T. Teoretisk kan ei slik sameint behandling vere å føretrekke; dei strukturelle skilnadene dannar ikkje då aleine grunnlag for at ordet treng å dobbelspesifiserast i leksikon.

Sjølv om dobbelkonstruksjonen gjerne er for uformell til å bli inkludert i dei normerande ordbøkene, kan ein ikkje sjå forbi han når ein skal

konkludere teoretisk om statusen til *å*. Han bidrar dessutan til ei meir einskapleg teoretisk forståing; den «offisielle» kategorien subjunksjon dekker ikkje ordets mest grunnleggande eigenskapar. I botnen er eg samd med Faarlund (2015) i at kopieringsteorien dekker den moderne norske *å*-empirien – òg når vi inkluderer kontrafaktivane, som Faarlund ikkje omtalar. Ubøyelege småord utan etablert ordklassetilhøyrse kallar vi *partiklar* (jf. tyd. 2 av *partikkel* i NOB), og sidan *å* er assosiert med eit generelt infinitt trekk i T, er den deskriptivt best eigna nemninga *infinittheitspartikkel*. Vi kan gå ut frå at infinittheitstrekket òg blir sjekka i *å*-konstruksjonar utan verb, jf. (7) lenger oppe.

Eit alternativ er likevel å forstå *å* som prinsipielt heterosemisk, og at ordet kan ha ulik status i ulike konstruksjonar og på tvers av dialektsystem. I så fall er det ei større empirisk oppgåve å kartlegge kva som er godtakande av meir marginale konstruksjonar i ulike talemål.

Ein finn òg tilfelle med setningsadverbial både før og etter *å*, jf. (19) frå *Leksikografisk bokmålskorpus* (Fjeld, Nøklestad & Hagen 2020).<sup>16</sup>

- 19) Sjefene på skips- og luftfartøyer er for eksempel pålagt **ikke** å **bevisst** forstyrre sambandssystemene for motparten.

Dette er mogleg å gjere greie for ved å gå ut frå at *bevisst* er adjungert til verbfrasen (VP) og *ikke* til tempusfrasen (TP) (jf. strukturane i (11) lenger oppe). Dersom ein skal argumentere for at *å* likevel står i C i (19), må ein godta at *å*-infinitivar skil seg prinsipielt frå *at*-setningar ved at førstnemnde *kan* ha nokre adverbial adjungert til subjunksjonsfrasen (CP). I så fall er CP-en i *å*- vs. *at*-setningar ulik på fleire måtar enn med omsyn til finittheit. Eg hallar sjølv mot at *å* er basegenerert i T, slik Faarlund (2015) hevdar. Då er *å* først og fremst assosiert med infinittheit i T, men *kan* innta C-posisjonen ved kopiering. Å markere infinittheit er altså *å* sin G-semantiske eigenskap, uavhengig av om ordet inntar ein syntaktisk posisjon som subjunksjon eller ikkje. Moglegheita til å «bli subjunksjon» gjer nettopp at ordet kan kategoriserast som heterosemisk, men utan at det rokkar ved den ordklassenøytrale nemninga infinittheitspartikkel.

Prinsipielt er det då ikkje noko i vegen for at *å* si ordklassetilhøyrse er ulik i ulike dialektsystem, t.d. at ordet oftare/lettare kan realiserast på subjunksjonsplassen i nokre dialektar enn andre. Ein skriftspråkleg kon-

16. Takk til eine fagfellen, som bringa dette dømet på bane.



sekvens bør i så fall vere at all slik variasjon er godtakande grammatikk i nynorsk òg. Dersom dialektgrammatikkar skal spele ei rolle i nynorsk, men t.d. i mindre grad i bokmål, bør ein nettopp kunne opne for at eit ord kan vere heterosemisk i nynorsk utan at det med naudsyn er det i bokmål. Slik variasjon bør dermed òg inkluderast i normative grammatikkar og ordbøker.

Å generalisere infinittheitspartikkelen til éin av dei ti konvensjonaliserte ordklassane er ikkje noko vidare poeng i, slik det heller ikkje er for predikasjonspartikkelen *som* (jf. Halmøy & Berg-Olsen 2012: 17–18). Det viktigaste vitskaplege poenget må vere å generalisere på empirisk mest mogleg haldbar måte (altså som infinittheitspartikkel i tilfellet *å*). I neste del blir det foreslått korleis slike generaliseringar kan formidlast i leksikografien.

## 5. Overordna ordbokproblem

Som nemnt i 3.2 har preposisjonsartiklane i NO særleg i 2014-prosjektet (dvs. etter 2002) gått i retning av å få ein «kognitivt» fundert struktur i tråd med den Runde & Kristoffersen (2006) fremjar. Den typiske disposisjonen i preposisjonsartiklane er at dei først har ein lokativ bolk (A), ein temporal bolk (B) og deretter ulik abstrakt bruk (C), før partikkel-samband følger anten seinare i artikkelen eller i ein separat adverbartikkel (pga. gamal ordklasseinndeling). Dette gjeld òg dei ovannemnde preposisjonsartiklane frå NO, band 7 og 8, sjølv det er forfekta eit litt anna syn i artiklane av Aa nemnde i 3.2. Skulle ein følge fullt ut synet i desse artiklane, burde ein bruke den første overordna tydinga til å presentere den best eigna samla kategorien og G-semantikken. Det ville krevje eit nivå over stor bokstav A, B, C osv., som er S-semantiske kategoriar. Det næraste føredømet ein kjem ein slik struktur i leksikografisk funksjonsordformidling i dag, er nettopp av *II som* i NO, band 10. For *å* sitt vedkomande må det innebere at rolla som infinittheitspartikkel blir løfta opp og fram i ordbokoppslaget. Det er det som verkar å gjennomsyre all bruken av *å*, sjølv om den syntaktiske distribusjonen (og såleis kanskje ordklassane) varierer.

Ein ordbokartikkel som er teoretisk strukturert «ovanfrå og ned» (jf. Runde & Kristoffersen 2006), vil gi det beste oversynet over bruksmangfaldet så lenge G-semantikken er godt postulert og S-semantikken er ka-

tegorisert i «rimelege» bolkar. Ein reint materialstyrt «nedanfrå og opp»-struktur får ikkje formidla kva som er meir grunnleggande enn noko anna, og vil med naudsyn òg bli ein artikkel som er vanskeleg å forhalde seg til som brukar. Det er berre å slå opp på preposisjonsoppslag i dei første NO-banda (t.d. *av* og *for*) for å få stadfesta dette. Motsett kan ei ulempe med den teoretisk drivne strukturen vere om ein lagar for rigide bolkar, slik at det blir påfallande og mange empiriske døme som ikkje passar heilt inn, og som anten blir fremja som spesielle «undertydingar» eller (bevisst) oversette. Ei normerande ordbok må likevel sjå forbi ein heil del av bruken, og då er den teoridrivne formidlinga tvillaust den ryddigaste (jf. Aa 2013: 157; 2015a: 122). All S-semantikken kan ikkje ein-gong dokumentasjonsordbøker gjere greie for. Derfor er det overordna å formidle ein tydeleg og predikerande struktur. For dei normerande ordbøkene, som i større grad enn dokumentasjonsordbøkene konsentrerer seg om hovudmønster (og der bruksfrekvens spelar ei endå større rolle), er ein predikerande struktur desto viktigare. Med det vil ein legge føringar for det som ikkje er eksplisitt sagt (jf. Aa 2013: 151–152). Funksjonsordartiklane får på denne måten nokre likskapstrekk med t.d. føreleddsartiklar gjennom at dei er «opne» og har innbaka unemnd, men (vonleg) predikerbar empiri. Å formidle ein open og predikerande struktur til både ekspertar og lekfolk er ei stor oppgåve som set krav til leksikografen. Det kan fort vere ei pedagogisk fallgruve å legge *for* stor vekt på ytterkantane òg, som kan både villeie og forvirre. Men eg trur at ein kan gjere lurt i å vise fram noko av den meir marginale grammatikken anten litt perifert i ordartikkelen eller gjennom undertydingar (eller liknande) i dei mest nærliggande tydingsbolkane.

Kva med ordklasseføring? Funksjonsorda er det ikkje alltid greitt å plassere i éin ordklassekategori, slik Halmøy & Berg-Olsen (2012) syner med *som*. *Som* er kategorisert som konjunksjon i NO, men er sjølv sagt mykje brukt som relativsubjunksjon òg (men subjunksjonar er ikkje ein eigen ordklasse i NO). Kanskje kunne ein ønske seg ei revitalisering av ei gamal løysing som ein brukte tidlegare, nemleg å liste opp alle ordklassane ordet tangerer, fremst i oppslaget. Til oppslaget *IX for* i NO, band 3, står det «prep, adv, konj». Ein måte å kategorisere på er gjennom ordklassetilhørsle, men for funksjonsorda kan det vere meir hemmande enn oppklarande. Og *å* er altså eit tilfelle der verken den tidlegare (infinitivmerke) eller nyare (subjunksjon) ordklassen dekkjer heile bruken. Viss ein avgrensar seg til å føre opp éin ordklasse, vil denne gi nokre

galne føringar uansett. Men ei overordna G-semantisk skildring ville i det minste bøtt på skaden.

## 6. Avrunding

Denne artikkelen har hatt to formål: å kaste lys over empiri som fortel oss om statusen til «infinitivmerket» *å* og å diskutere korleis dette og liknande grammatiske problem (ved funksjonsord) kan formidlast i ordbøker. Eg har særleg lagt vekt på synkrone *å*-data (både skriftlege og munnlege), men inkludert diakrone perspektiv. *Å* kan stå både på subjunksjonsplassen (C) og i setningas «midtfelt» (T), og det kan innleie meir enn infinitivar. Verken «subjunksjon» eller «infinitivmerke» er derfor heilt dekkande for all bruken av ordet i dag. Eg har i staden foreslått den ordklassenøytrale nemninga *infinittheitspartikkel*.

Parallelt med at *å*-forskinga har skote fart i dette tusenåret, har grenselandet mellom lingvistik og leksikografi òg vorte utforska. Medan prototypisk bruk er viktig å formidle i leksikografien, spelar marginale konstruksjonar ei viktig rolle i grammatikken. Eg har argumentert for at òg sistnemnde har ein viktig plass i den grammatiske leksikografien, og ikkje berre kan avfeiast som normativt feil. I formidlinga av funksjonsord må leksikografien vere språketeoretisk fundert, og dei relevante ordbokartiklane bør ha ein predikerande struktur (både av teoretiske og formidlingsmessige grunnar). Etter mitt skjøn er den beste måten å gjere det på å postulere ein G-semantikk som gjennomsyrrer all bruken av ordet (S-semantikken). Den mest grunnleggande eigenskapen til *å* i denne samanhengen er å markere generell infinittheit.

## Litteratur

- Adger, David. 2003. *Core syntax. A minimalist approach*. Oxford: Oxford University Press.
- Askedal, John Ole. 2016. Anmeldelse av *Norsk Ordbok* bind 12 – med noen betraktninger omkring verket i sin helhet. *Lexico-Nordica* 23, 185–205.

- Bjorvand, Harald, Even Hovdhaugen & Hanne Gram Simonsen (red.). 1982. *Språkvitenskap. En elementær innføring* (2. utg.). Oslo: Universitetsforlaget.
- Bouchard, Denis. 1995. *The semantics of syntax. A minimalist approach to grammar*. Chicago: The University of Chicago Press.
- . 2002. *Adjectives, number and interfaces. Why languages vary*. Oxford: Elsevier Science.
- Bowers, John. 1993. The Syntax of Predication. *Linguistic Inquiry* 24, 591–656.
- Bye, Patrick, Trond Trosterud & Øystein Vangsnes. 2003. *Språk og språkvitenskap. Ei innføring i lingvistikk*. Oslo: Samlaget.
- Dikken, Marcel den. 1995. *Particles. On the syntax of verb-particle, triadic and causative constructions*. Oxford: Oxford University Press.
- Eide, Kristin Melum. 2005. *Norwegian modals*. Berlin: Walter de Gruyter.
- . 2011. The ghost of the Old Norse subjunctive: the Norwegian subjunctive participle. *Groninger Arbeiten zur germanistischen Linguistic*, 52(3), 1–28.
- Fanselow, Gisbert, Caroline Féry, Ralf Vogel & Matthias Schesewsky. 2008. Gradience in Grammar. I: *Gradience in Grammar. Generative Perspectives*, red. Gisbert Fanselow, Caroline Féry, Ralf Vogel & Matthias Schesewsky, 1–21. Oxford: Oxford University Press.
- Fjeld, Ruth Vatvedt, Anders Nøklestad & Kristin Hagen. 2020. Leksikografisk bokmålskorpus (LBK) – bakgrunn og bruk. I *Leksikografi og korpus. En hyllest til Ruth Vatvedt Fjeld, Oslo Studies in Language* 11(1), red. Janne Bondi Johannessen & Kristin Hagen, 47–59.
- Faarlund, Jan Terje. 1988. *Morfologi. Bøyingssystemet i nynorsk og bokmål*. Oslo: Samlaget.
- . 1995. *Morfologi. Bøyingsystemet i nynorsk og bokmål* (2. utg.). Oslo: Samlaget.
- . 2003. Reanalyse og grammatikalisering i norske infinitivkonstruksjoner. I: *Språk i endring. Indre norsk språkhistorie*, red. Jan Terje Faarlund, 57–79. Oslo: Novus forlag.
- . 2005. Grammatikalisering og degrammatikalisering i norske infinitivkonstruksjoner. I: *Grammatikalisering og struktur*, red. Lars Heltoft, Jens Nørgård-Sørensen & Lene Schøsler, 85–100. København: Museum Tusulanums Forlag.

- .2010. [Melding av] Norsk Ordbok, band VIII. *Lexico Nordica* 17, 313–320.
- . 2015. The Norwegian infinitive marker. *Working Papers in Scandinavian Syntax* 95, 1–10.
- . 2019. *The syntax of Mainland Scandinavian*. Oxford: Oxford University Press.
- Faarlund, Jan Terje, Svein Lie & Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Garbacz, Piotr. 2010. *Word order in Övdalian: a study in variation and change*. Doktoravhandling, Lunds universitet.
- Gram Simonsen, Hanne, Rolf Theil Endresen & Even Hovdhaugen (red.). 1988. *Språkvitenskap. En elementær innføring* (3. utg.). Oslo: Universitetsforlaget.
- Halmøy, Madeleine & Sturla Berg-Olsen. 2012. I leksikografiens periferi – tydingsbeskriving av *som* i *Norsk Ordbok*. *Lexico-Nordica* 19, 17–38.
- Holmberg, Anders & Christer Platzack. 1995. *The role of inflection in Scandinavian syntax*. Oxford: Oxford University Press.
- Hovdhaugen, Even (red.). 1976. *Språkvitenskap. En elementær innføring* (1. utg.). Oslo: Universitetsforlaget.
- Jackendoff, Ray. 1983. *Semantics and cognition*. Cambridge, MA: MIT Press.
- Kayne, Richard. 1985. Principles of particle constructions. I: *Grammatical Representation*, red. Jacqueline Guéron, Hans-Georg Obenhauer & Jean-Yves Pollock, 101–140. Dordrecht: Foris.
- Kinn, Torodd. 2007. Den historiske utviklinga til ord på *-vis*. *Maal og Minne* 2 (2007), 158–186.
- Langacker, Ronald. 1987. *Foundations of cognitive grammar: Theoretical prerequisites (vol. 1)*. Stanford: Stanford University Press.
- Larsen, Darrell. 2014. *Particles and particle-verb constructions in English and other Germanic languages*. Doktoravhandling, University of Delaware.
- Lichtenberk, Frantisek. 1991. Semantic change and heterosemy in grammaticalization. *Language* 67, 475–509.
- Myrvoll, Klaus Johan. 2005. [Melding av] *Språk i endring. Indre norsk språkhistorie*, red. Jan Terje Faarlund. *Norsk Lingvistisk Tidsskrift* 23 (2), 271–281.

- Mørck, Endre. 2019. Infinitivsmerkets status i mellomnorske kontrollkonstruksjoner. *Maal og Minne* 2 (2019), 101–144.
- Newmeyer, Frederick. 1998. *Language form and language function*. Cambridge, MA: MIT Press.
- . 2005. *Possible and probable languages. A generative perspective on linguistic theory*. Oxford: Oxford University Press.
- Nordgård, Torbjørn & Tor Anders Åfarli. 1990. *Generativ syntaks. Ei innføring via norsk*. Oslo: Novus.
- Norsk språkråd. 2005. Grammatiske termer til bruk i skoleverket. Tilråding fra Norsk språkråd og Utdanningsdirektoratet. Lasta ned frå <https://www.sprakradet.no/localfiles/gramterm.pdf>
- Ommeren, Rikke van. 2010. «Ja, jøi la an på å tåla oppdaling, ja»: En sociolinguistisk studie av språklige praksisformer blant voksne innvandrere i Oppdal. Masteroppgåve, NTNU.
- Runde, Ålov & Kristian Emil Kristoffersen. 2006. Preposisjonen til mellom lingvistisk teori og leksikografisk praksis. *Norsk lingvistisk tidskrift* 24 (1), 3–25.
- Rønhovd, Jarle. 1993. *Norsk morfologi*. Oslo: Gyldendal.
- Sandøy, Helge. 1991. Attraksjon av supinum i færøysk og norsk. *Danske folkemål* 33, 251–262.
- Schütze, Carson. 2010. Linguistic evidence and grammatical theory. *WIREs Cognitive Science* (2), 206–221.
- Teleman, Ulf. 2004. [Melding av] *Språk i endring. Indre norsk språkhistorie*, red. Jan Terje Faarlund. *Maal og Minne* 2004, 122–126.
- Theil Endresen, Rolf. 1988. Morfologi. I: *Språkvitenskap. En elementær innføring* (3. utg.), red. Hanne Gram Simonsen, Rolf Theil Endresen & Even Hovdhaugen, 76–140. Oslo: Universitetsforlaget.
- Wetås, Åse & Leiv Inge Aa. 2017. Museale betraktningar av eit moderne ordbokverk – tilsvar til John Ole Askedal si melding av *Norsk Ordbok*. *Lexico-Nordica* 24, 289–300.
- Aa, Leiv Inge. 2011. Partiklar inn i ordbøkene? Om tradisjonell og nyare kategorisering av preposisjonar. *Lexico Nordica* 18, 149–166.
- . 2013. Er ordbøker lingvistiske nok? Om preposisjonssemantikk og ein grammatisk basis for ordboksartiklar. *Lexico Nordica* 20, 145–163.
- . 2014. Å ikkje akte infinitivsmerket. I: *Frå A til Å. Veneskrift til Dagfinn Worren på 70-årsdagen 28. desember 2014*, red. Olaf Almenningen, Oddrun Grønvik, Helene Urdland Karlsen, Andreas Tandberg, Lars S. Vikør & Åse Wetås, 135–138: Oslo: Novus.

- . 2015a. Mangfold og einskap med preposisjonar – og vitskaplege krav til moderne ordbøker. I : *Ivar Aasen ute og heime – om moderne språkdokumentasjon etter Ivar Aasen*, red. Jan Ragnar Hagland & Åse Wetås, 110–125. Trondheim: *Det Kongelige Norske Videnskabers Selskabs Skrifter* 1/2015.
- . 2015b. Metodologiske og teoretiske problem for generative dialektologar. *Norsk Lingvistisk Tidsskrift* 33 (2), 121–128.
- . 2017a. Godt mot, Olaf! I: *I teneste for nynorsken. Heidersskrift for Olaf Almenningen 70 år*, red. Knut E. Karlsen, Dagfinn Rødningen & Håvard Tangen, 233–244. Oslo: Novus.
- . 2017b. Går artikkelen opp? I: *Målreising og morsmålsdokumentasjon. Heidersskrift til Oddrun Grønvik ved 70-årsleitet*, red. Bjørghild Kjelsvik, Christian-Emil S. Ore, Lars S. Vikør, Åse Wetås & Dagfinn Worren, 173–186. Oslo: Novus.
- . 2018. Preposisjonen og partikkelen *med* som samanstillar. *Norsk Lingvistisk Tidsskrift* 36 (1), 139–159.
- . 2020. *Norwegian verb particles*. Amsterdam: John Benjamins.
- Åfarli, Tor Anders. 1997. *Syntaks. Setningsbygning i norsk*. Oslo: Samlaget.
- . 2019. Syntaks: overordna leddstillingsvariasjon. I: *Språket som system: norsk språkstruktur*, red. Heidi Brøseth, Kristin Melum Eide & Tor Anders Åfarli, 203–230. Bergen: Fagbokforlaget.
- Åfarli, Tor Anders & Kristin Melum Eide. 2001. Predication at the Interface. *ZAS Papers in Linguistics* 26, 35–59. Doi: <https://doi.org/10.21248/zaspil.26.2001.137>
- . 2003. *Norsk generativ syntaks*. Oslo: Novus.
- Aasen, Ivar. 1864. *Norsk Grammatik*. Christiania: Mallings.

## Ordbøker og andre empiriske kjelder

- NO = *Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet*. <http://no2014.uib.no>
- . 1966. Band 1, hovudred. Alf Hellevik. Oslo: Samlaget.
  - . 1989. Band 3, hefte 1. Oslo: Samlaget.
  - . 1994. Band 3. Oslo: Samlaget.
  - . 2008. Band 7, hovudred. Oddrun Grønvik, Laurits Killingbergtrø, Lars S. Vikør & Dagfinn Worren. Oslo: Samlaget.

- . 2009. Band 8, hovudred. Oddrun Grønvik, Lars S. Vikør & Dagfinn Worren. Oslo: Samlaget.
  - . 2012. Band 10, hovudred. Oddrun Grønvik, Lars S. Vikør & Dagfinn Worren. Oslo: Samlaget.
  - . 2016. Band 12, hovudred. Oddrun Grønvik, Helge Gundersen, Lars S. Vikør & Dagfinn Worren. Oslo: Samlaget.
- BOB = *Bokmålsordboka*. Henta frå <https://ordbok.uib.no>  
NOB = *Nynorskordboka*. Henta frå <https://ordbok.uib.no>  
*Nynorskorpuset*. <http://no2014.uib.no/korpuset>  
SA = *Setelarkivet til Norsk Ordbok*. Henta frå <http://usd.uib.no/perl/search/search.cgi?tabid=436&appid=8>  
Aasen, Ivar. 1850. *Ordbog over det norske Folkesprog*. Kristiania: Werner & Comp.  
–. 1873. *Norsk Ordbog med dansk Forklaring*. Christiania: Malling.

## English abstract

The article has two main purposes: to map the theoretical status of the Norwegian «infinitive marker» *å* ‘to’ and to discuss the inclusion of linguistic theories in dictionary entries. Over the last 20 years, *å* ‘to’ has been investigated thoroughly diachronically and synchronically. There is some consensus about its change of status from complementizer in Old Norse to verbal clitic in early Modern Norwegian and back to complementizer in later Modern Norwegian. The synchronic complementizer status is criticized here, and a lower infinite particle is suggested. The last 15 years, linguistic theories have also entered Norwegian lexicography and made their impact on functional word entries. Principles for the lexicographic presentation of *å* ‘to’ and other functional words are discussed and evaluated. Essentially, a dictionary entry which is accurate enough, but adequately open, is pursued, in order to predict empirical facts which are not explicitly mentioned in the dictionary.

Leiv Inge Aa  
Noregs teknisk-naturvitskapelege universitet  
Institutt for lærarutdanning  
NO-7491 Trondheim  
[leiv.inge.aa@ntnu.no](mailto:leiv.inge.aa@ntnu.no)



## Innkalling til årsmøte i Bymålslaget

Torsdag 16. februar 2023 kl. 1800  
4. etg. Fysikkbygningen, Ø465  
Universitetet i Oslo, Blindern

Dagsorden:

1. Godkjenning av innkalling og dagsorden
2. Årsrapport for 2022
3. Regnskap for 2022
4. Rapport fra redaksjonen i Maal og Minne
5. Drøfting av innkomne forslag
6. Drøfting av virksomheten i tida framover
7. Valg

Framlegg til saker som skal tas opp på årsmøtet, må være sendt til lederen av laget seinest tre uker før møtedagen. Framlegg kan sendes med post til Andreas Haraldsrud, Tjyruhjellveien 29, 3512 Hønefoss, eller på e-post til [a\\_haraldsrud@hotmail.com](mailto:a_haraldsrud@hotmail.com).

Inngangsdøra til bygningen er låst etter klokka 16:00. Møt derfor opp 17:50 utafor hovedinngangen, så vil du bli sluppet inn.

Vel møtt!

Andreas Drolsum Haraldsrud

Leder