

Om artiklene i denne boka og *Nordisk dialektkorpus*

Janne Bondi Johannessen og Kristin Hagen

Innledning

Det er ikke ofte det presenteres mye ny kunnskap på én gang om moderne dialekter og talespråk. Denne boka tror vi derfor vil glede et større publikum, på samme måte som vi selv har hatt mange gode opplevelser i arbeidet med den.

Selve boka springer ut av et seminar om nordiske dialekter, språk og talemålsforskning som vi arrangerte i 2013.¹ Foredragsholderne var toppforskere fra Norge og andre nordiske land, og foredragene spente vidt, både når det gjaldt språkvariantene som var omtalt, og de språklige emnene – fra lydlige fenomener til bruk av ulike ord og større konstruksjoner. Bakgrunnen for seminaret var at et stort empirisk materiale om talespråk i Norden nylig var ferdigstilt, nemlig *Nordisk dialektkorpus* (som vi beskriver litt senere). Foredragsholderne var bedt om å bruke dette korpuset, og alle foredragene inneholdt derfor helt ny forskning på et helt ferskt materiale. Denne boka presenterer en del av dette.

Forfatterne er blitt bedt om å gjøre artiklene tilgjengelige for et noe bredere publikum enn det de vanligvis skriver for. Det innebærer at de her treffer lesere som ikke venter seg høyttravende teori-diskusjoner, men heller framlegging av data som ikke har vært all-

1. Seminaret het *Dialektseminar 2013 – Ny forskning basert på Nordisk dialektkorpus* og fant sted på Universitetet i Oslo 10.–11. januar 2013. <http://www.hf.uio.no/iln/om/organisasjon/tekstlab/aktuelt/arrangementer/2012/nordisk-dialektforskning.html>

ment kjent, gjerne fordi det til i dag ikke har vært empirisk materiale enkelt tilgjengelig for denne typen forskning. Vi får for eksempel vite at det er geografiske forskjeller mellom bruken av ord som *mor* og *mamma*, og at de skandinaviske språkene i likhet med stillehavsspråk har ulikt grammatisk uttrykk for eierskap alt etter hvem som eier hva, slik at man uttrykker eierskapet til kroppsdeler på en annen måte enn eierskapet til trær.

Artiklene handler om talespråk i vid forstand. Det betyr at noen av dem tar for seg språket slik det snakkes i hele landet, eller i flere land, uten å skille mellom folk på ulike steder eller i forskjellige aldre, mens det i andre artikler er lagt vekt på nettopp å studere språklige forskjeller enten de skyldes geografi eller andre ting. Nedenfor gir vi en kort beskrivelse av de enkelte artiklene i boka. Deretter presenterer vi *Nordisk dialektkorpus*. Til slutt i dette kapitlet gir vi en oversikt over fagfeller som har lest gjennom artiklene og foreslått endringer. Denne prosessen har vært viktig for det endelige resultatet.

Om de forskjellige artiklene

Nedenfor gir vi en kort oversikt over artiklene, inndelt etter hva de handler om.

Syntaks: Ordstilling og konstruksjoner

Kristine Bentzen har i sin artikkel *Setninger med 'kanskje' i forfeltet i skandinavisk, med vekt på norsk og svensk* undersøkt den typen setninger som angis i tittelen, og oppdaget en rekke nye sider ved språkene våre. For det første ser det ut til å være stor forskjell mellom skriftspråk og talespråk når det gjelder ordstilling og ordet *kanskje*, for det andre er det en klar forskjell mellom norsk og svensk talespråk på dette punktet. Begge talespråkene er også forskjellige fra sine respektive skriftspråk, der *kanskje* gjerne etterfølges av verb. Norsk talespråk har typisk ordet *kanskje* først, fulgt av sub-

jekt (*kanskje jeg tok feil*), mens svensk gjerne har dette ordet etter subjektet (*du kanskje har lite frågor*).

Tor Arne Haugen undersøker formen på adjektivets komplement i norske dialekter i artikkelen *Adjektiv med utfyllingar: Geografisk og medial variasjon*. Han gjør to nye funn. Det ene er at det er stor forskjell på skriftspråk og talespråk, nemlig at det er en mulig type adjektivutfylling som aldri forekommer i tale (*Hun var ham kjær*). Det andre er at han får stadfestet empirisk en antagelse grammatikere tidligere har gjort: at det går et klart dialektskille når det gjelder om utfyllingen har med preposisjon eller ikke: *flink til å sykle* eller *flink å sykle*. Den første typen finnes i østnorsk (inkludert trøndersk) og den andre i vestnorsk (inkludert nordnorsk).

Janne Bondi Johannessen, Marit Julien og Helge Lødrup har skrevet artikkelen *Preposisjoner og eierskapsrelasjoner: et menneskesentrert hierarki*. Mens det har vært kjent siden begynnelsen av forrige århundre at stillehavsspråk skiller grammatisk mellom eierskap til ting som er nært knyttet til mennesket (som kroppsdeler og slektskapsrelasjoner), og andre ting (som trær eller kopper), har det helt fram til vår tid vært mindre kjent at de skandinaviske språkene har det samme skillet. De tre forfatterne viser at det er mange spesielle uttrykksmåter for de eierskapsrelasjonene som er nærmest mennesket, mens det er færre for de fjernere. For eksempel er det greit å unnlate å ha bestemt form av substantiv ved slektskap: *Den fikk jeg kjøpt av bror til Alfhild*. Men man ikke kan si **Jeg beundret blomst til Anna*. Forfatterne viser at det er en rekke grammatiske forskjeller avhengig av hvem som eier hva. De viser også at det er store forskjeller de nordiske språkene imellom.

Ida Larsson og Björn Lundquist ser i sin artikkel *Objektsplacering vid partikelverb i norska dialekter och äldre svenska* nærmere på et fenomen som kanskje er et av de syntaktiske hovedskillene mellom norsk og svensk, nemlig partikkelplassering. Den typiske forskjellen mellom språkene illustreres i disse to setningene: *Vi kastet den*

ut vs. *Vi kastade ut den*. Utviklingen i svensk har vært stor, fra muligheten til å ha både leksikalske og pronominale objekter foran partikkelen, via en periode med bare pronomen foran partikkelen og til dagens situasjon, der ingen type objekt kan komme foran partikkelen. I norsk varierer det derimot geografisk om trykksvake objektspronomen er plassert foran eller etter partikkelen.

Øystein Alexander Vangsnes og **Marit Westergaard** undersøker ordstilling i artikkelen *Ka korpuse fortæll? Om ordstilling i hv-spørsmål i norske dialekter*. De har sett nærmere på spørresetninger av typen *Ka du synes om elgjakta?* Det er kjent at mange norske dialekter har slike setninger, der altså verbet ikke står direkte etter spørreordet, slik det skal i skriftspråket (*Hva synes du om elgjakta?*). Vangsnes og Westergaard har undersøkt forholdet mellom denne typen ordstilling og ulike typer spørreord og spørreuttrykk. De finner at lange spørreuttrykk (som *Hvor mye kollektivtrafikk*) sjeldnere opptrer sammen med denne ordstillingen enn korte spørreord (som *Hva*). Likevel er det geografiske forskjeller her, og det er Nord-Norge som har de strengeste reglene, mens Nordvestlandet og Rogaland tillater denne ordstillingen for alle slags spørreuttrykk. De finner ellers at selve ordstillingen er utbredt over store deler av Norge, inkludert Østlandet.

Morfologi og fonologi

Lars-Olof Delsing tar i artikkelen *Verbsystemet i en västerbottensk dialekt* utgangspunkt i en nyere lingvistisk hypotese om at det tradisjonelle germanske verbsystemet kan deles i fire på bakgrunn av hvilken bøyning det enkelte verbet har. Denne firedelingen er basert på at hvert verb enten er bøyd i fortid eller ikke (tempuskategorien), som *kjøpte* vs. *kjøper*, og at det enten kan bære en hel setning alene eller ikke (finitthetskategorien), som *kjøpte* vs. *kjøpt*. Hypotesen sier at engelsk, med sitt forenklete morfologiske system, har mistet kategorien finitthet. Delsing undersøker dialekten i Lycksele lappmark i Västerbotten. Den har gjennomgått mange endringer,

og Delsing viser at disse endringene systematisk lar seg forklare innenfor det samme perspektivet, altså med tap av finittet. På grunn av forskjellige lydutviklinger i dialekten er det stort sett bare skillet mellom fortid og nåtid som gjenstår, slik at infinitiv og presens på den ene siden har samme form, f.eks. *tjöp* 'kjøpe' og *kjøper*, og preteritum og supinum på den andre siden *tjöfft* 'kjøpte' og 'kjøpt'.

Thórhallur Eythórsson og **Jóhannes Gísli Jónsson** sammenligner islandsk og færøysk i sin artikkel *Oblike subjekter i færøysk og islandsk*. Det er kjent at disse øynordiske språkene kan ha subjekt i akkusativ og dativ, det sistnevnte illustrert i henholdsvis færøysk *Henni dámar føroyskan tónleik* ('Hun liker færøysk musikk') og islandsk *Mér leiðist að tala norsku* ('Jeg er lei av å snakke norsk'). Utbredelsen har vært gjenstand for flere språklige undersøkelser. Dativ tar i stadig større grad over for de tradisjonelle akkusativsubjektene i islandsk, og fenomenet har fått et eget kallenavn – dativsyken – som sier mye om at fenomenet er kjent og utbredt. I færøysk tyder utviklingen på at dativsubjektene forsvinner til fordel for nominativ. Eythórsson og Jónsson finner at korpuset dessverre fortsatt har for lite færøysk og islandsk materiale til at fenomenet kan studeres i sin fulle bredde.

Unn Røyneland og **Brit Mæhlum** har kalt sin artikkel *Det norske dialektlandskapet: sosiolingvistiske perspektiv på bruk av Nordisk dialektkorpus*. De tar et prinsipielt perspektiv på metodeforskjellene ved dialektinnsamling i tradisjonell dialektologi og i sosiolingvistik. I den førstnevnte er man interessert i å finne det som forener innenfor et geografisk avgrenset område, mens i den andre er man opptatt av det som skiller sosiale grupper innenfor et slikt område. De kaster videre et kritisk blikk på nytten av dialektkart, og advarer mot å stole for mye på dem. Likevel finner de at dialektkorpuset kan brukes til å se på den geografiske utbredelsen av retrofleks, og undersøker her retroflektoring med /-s/ og med /-t/ i to dialekter i Telemark og på Sunnmøre som begge tradisjonelt ikke har hatt

retrofleksjer. I begge dialektene viser det seg at det har vært en klar utvikling i retning av mer retrofleksjering, selv om det også er noen forskjeller mellom dem.

Leiv Inge Aa, Kristin Melum Eide og Tor Anders Åfarli undersøker i artikkelen *Somme mytar er fort avklede: perfektum partisipp i dialektkorpuset* perfektum partisipp-forma, og presenterer tre ulike funn. Først undersøkes partisippkongruens der det formelle subjektet er enten *det* eller *der*. Det har vært vanlig å anta at partisippet samsvarsbøyes med det formelle subjektet *det*, mens det er det leksikalske subjektet som bestemmer samsvarsbøyingen når det formelle subjektet er *der*. Korpusdataene viser derimot noe annet, nemlig at slik kongruens godt kan skje, men bare i dialekter som også har *der* som mulig formelt subjekt. Man finner altså helt uventet eksempler med kongruens selv om det formelle subjektet er *det*, som i *Det vart skotne mange elgar i fjor* i stedet for, som ventet, uten kongruens, som i *Det vart skote mange elgar i fjor*.

En annen vanlig antagelse er at partikkelinkorporering i partisippet samvarierer med muligheten for at partisippet kan kongruere. Men forfatterne finner eksempler som *Butikken ble nedlagt*. Dersom den vanlige antagelsen hadde slått til her, burde partisippet med partikkelen vært bøyd i hankjønn i samsvar med hankjønnssubjektet: *nedlagd*. I artikkelen vises det altså at dette er en myte, og at det ene godt kan eksistere uten det andre. Til slutt diskuterer forfatterne påstanden om at konjunktiv er en kategori som døde i norsk for et halvt årtusen siden. Påstanden tilbakevises kontant, og de viser at konjunktiv som irrealismarkør fortsatt eksisterer over hele landet, og at det gjerne er partisippformen som brukes. Fra Oppdal er det for eksempel opptak av en ung mann som bruker partisippform når han snakker om noe som ikke har skjedd, *de kunna vora arrti*, og preteritumsform ellers ved vanlig modal betydning: *åss fann ut att åss kunn jora sånn*.

Leksikografi: leksikalske ord, partikler og funksjonsord

Ruth Vatvedt Fjeld har et leksikografisk perspektiv i artikkelen *Mor eller mamma? Om dialektal variasjon og endring i sentrale norske slektskapstermer i norske dialekter*. Hun tar for seg norske slektskapstermer, gir en oversikt over det leksikalske inventaret som norsk har på dette feltet, og analyserer variasjonen mellom alternative former geografisk og (tentativt) diakront – altså over tid – primært på bakgrunn av data fra Nordisk dialektkorpus og Leksikografisk bokmålskorpus. Hun har funnet at noen slektskapstermer, som *mamma* og *pappa*, er blitt mye vanligere nå enn de var tidligere, mens andre taper terreng. Det er også geografiske forskjeller – *mor* og *far* er mer vanlig på Vestlandet og Østlandet, mens *mamma* og *pappa* er vanligere i Nord-Norge. Hun har også tatt for seg mange andre vanlige og uvanlige slektskapstermer.

Helge Sandøy skriver om mengdebestemmere i artikkelen *'Somme' har nok hørt 'eitkvart' om kvantoren 'nokon' før*. Han undersøker ordene *somme*, *nokon* og *einkvan*. Betydningen til disse ordene er i utgangspunktet svært ulik. For eksempel krever *somme* at mengden ikke utgjøres av en sammenhengende helhet. Det skal ikke være mulig å si *Ho satt heime somme dagar* dersom dagene fulgte rett etter hverandre. Sandøy benytter korpusets opptak fra forskjellige epoker til å undersøke mulige endringer i tid, og han ser også på den geografiske utbredelsen av disse ordene. Han finner store endringer for *somme*. De eldste opptakene (fra 1950- til 1970-tallet) gir ti ganger så mange treff, og fra hele Sør-Norge. Nå er det bare fjellområdene som har dette ordet i bruk. Sandøy presenterer også flere diskusjoner om årsakene til at kvantorene er ulikt fordelt i korpuset.

Jan Svennevig står bak artikkelen *"Det er nå trivelig der innpå, da" 'Nå' – en diskursmarkør i tilbakegang*. Her tar han for seg diskurspartikkelen *nå*, og viser at den har en rekke pragmatiske funksjoner. Den kan være en interjeksjon, som i svaret *Åh nå!* En annen

mulighet er at den kan være en partikkel med innrømmende eller konsessiv funksjon, som i *Nå har ikke vi idrettshall, da, men ...*, eller at den er en affektiv og epistemisk forsterker: *Det er nå trivelig der innpå, da*. Svennevig har i tillegg funnet ut at forsterkeren er i sterk tilbakegang, noe som er interessant fordi mye annen forskning stort sett har fokusert på partikler som er i vekst.

Åse Wetås sammenligner i artikkelen *Nordisk dialektkorpus og Norsk målføresynopsis – ei samanlikning av to kjelder til talemålsforskning frå kvar si tid* de to dialektsamlingene som er nevnt i artikkelens tittel, og som er fra henholdsvis 2010-tallet og 1950–1970. Hun gir en oversikt over likhetene og forskjellene på disse to språksamlingene, og viser at man godt kan lete etter de samme fenomenene i begge. I de tilfellene der et språklig trekk eller ord bare finnes i den eldre synopsisen, ser det ut som en språkendring har funnet sted. Ett eksempel blant flere hun trekker fram, er nektingsadverbet *itte*, som ble brukt i Lommedalen i Bærum ifølge den gamle målføresamlingen, synopsisen. Ingen av de mange korpustreffene på nektingsadverbet gir noe annet enn *ikke* for dette området i dag.

Om metode og *Nordisk dialektkorpus*

Når språkforskere skal undersøke et språk eller en dialekt, er det flere måter å gjøre det på. Når forskeren kan språket selv, er det ikke uvanlig at vedkommende bruker sin egen kunnskap, gjerne kalt introspeksjon. Hvis man kan språket dårligere, er det gjerne påkrevet å bruke andre enn seg selv for å framskaffe empirien, for eksempel ved å stille spørsmål til talere av språket eller lytte til opptak og lese tekster på språket. Imidlertid er alle disse metodene mangelfulle. Introspeksjon kan i verste fall føre til at forskeren lar seg farge av sin egen hypotese til å framskaffe data som passer med den. En mindre alvorlig følge er at forskeren kun finner det som vedkommende er på jakt etter, enten det dreier seg om ord eller om grammatiske konstruksjoner. Spørsmål til språkets talere gir heller

ikke så mange andre konstruksjoner enn dem man hadde i tankene, men svarene er man i det minste ikke fullstendig herre over selv som forsker.

Å bruke et fysisk språklig materiale, enten det er opptak eller tekster, har mange fordeler. Leter man etter én ting, kan man snuble over en annen, som kanskje er mindre kjent og mer interessant. Dataene ligger også der – utenfor forskeren selv – slik at de er mer objektive. En ulempe med å bruke datamateriale som tekster og opptak kan være at det er tungt og tidkrevende å arbeide med. Dessuten er det også tidkrevende og kostnadskrevenende om man skal framskaffe det selv. Man kan derfor knapt arbeide med mange dialekter eller studere mange taleres språk samtidig. Hvis man derimot har muligheten til å bruke et stort språkmateriale som inneholder både opptak og transkripsjoner av opptakene, som dekker et stort geografisk område med mange unge og gamle talere, og man kan søke direkte etter ord og konstruksjoner man er interessert i, og få ut lister av bare disse, med lydopptak av de samme utsnittene, ja, da er verden en annen. *Nordisk dialektkorpus* er et slikt materiale.

Innholdet i *Nordisk dialektkorpus*






Nordisk dialektkorpus (Johannessen et al. 2009) inneholder over 2,8 millioner løpeord med naturlig tale uttalt av 821 informanter fra 228 målepunkter i fem nordiske land.² Innsamling av materiale, transkripsjon og teknisk utvikling har pågått i over åtte år, og har involvert mange titalls forskere, assistenter og tekniske utviklere. Både dialektkorpuset og *Nordisk syntaksdatabase*³ (Lindstad et al.

2. Strengt tatt inneholder korpuset nøyaktig 2 824 585 ord pr. 27. mars 2014. Nytt materiale kan bli lagt til korpuset senere, og feil blir rettet fortløpende. Antall ord kan derfor endre seg noe.
3. Nordisk syntaksdatabase inneholder data fra spørreskjemaer med setninger som ble evaluert grammatisk av informantene for å kartlegge grammatisk variasjon i nordiske dialekter. Databasen er et viktig supplement til korpuset fordi spørreskjemaene inneholder konstruksjoner som ikke nødvendigvis er så frekvente i intervjuer og samtaler.

2009) er resultat av to store samarbeidssatsinger: nettverket ScanDiaSyn (Scandinavian Dialect Syntax, 2004–2009) og NORMS (Nordic Centre of Excellence in Microcomparative Syntax, 2005–2010). Nedenfor vil vi kort beskrive korpusmaterialet fra hvert enkelt land, før vi presenterer enkelte av mulighetene i *Nordisk dialektkorpus*.

Hvert deltakerland har hatt ansvar for finansiering, utvalg av informanter, innsamling og transkripsjon av eget materiale. Noe arbeid er gjort innenfor rammene av ScanDiaSyn og NORMS. Dette har ført til at korpusmaterialet ikke er helt ensartet, og at antall ord varierer fra land til land, slik tabell 1 viser. Alle optak har det til felles at de består av fri tale, og at denne talen er transkribert ortografisk på hvert enkelt lands standardspråk. Transkripsjonene er deretter automatisk grammatisk tagget. Det innebærer at hvert ord er merket med ordklasse og annen grammatisk informasjon, slik at språkforskeren kan søke direkte etter for eksempel *hive+verb perfektum partisipp*, og slik få fram alle forekomster på én gang, for eksempel *hivd, høvd, hevet, heve, hevve, bevi, beven, hevven, hiva*, fra hele Norge. Det går nesten ikke an å tenke seg hvordan dette kunne vært gjort før korpusets tid.

Vi kommer nærmere inn på hvert enkelt land nedenfor. For flere detaljer om korpuset, databasen og innsamlingen enn vi kan gi her, anbefaler vi interesserte å gå til nettsiden til *Nordisk dialektkorpus* (se referanselista sist i innledningskapitlet).

	Danmark 	Færøyene 	Island 	Norge 	Sverige 
Informanter	81	20	30	564	126
Målepunkter	15	5	7	163	38
Antall ord	211 266	62 411	56 001	2 187 046	307 861

Tabell 1: Antall informanter, målepunkter og ord for landene i Nordisk dialektkorpus. De norske tallene inkluderer materiale fra Norsk målførearkiv.

En kort oversikt over innholdet for de enkelte språkene gir vi her:

Dansk: Materialet fra Danmark er hovedsakelig samlet inn og transkribert av DanDiaSyn-prosjektet (2007–2008), et samarbeidsprosjekt mellom forskere ved Aarhus Universitet og Københavns Universitet. Materialet inneholder lydopptak av intervjuer med eldre informanter fra elleve målepunkter i Danmark. I tillegg kommer noen videosamtaler fra fire målepunkter tatt opp under en NORMS-workshop på Vestjylland i 2008. Både lyd- og videoopptak er ortografisk transkribert.

Færøysk: Opptakene fra Færøyene ble gjort under en NORMS-workshop på Færøyene i august 2008, og består av videoopptak av intervjuer og samtaler med eldre og yngre kvinner og menn. Alt materialet er ortografisk transkribert.

Islandsk: De islandske opptakene er av tre typer: Det nyeste materialet er lydopptak av samtaler fra 2013 med tjue informanter fra fem målepunkter på Island. Informantene er både eldre og yngre, kvinner og menn. Den andre typen er videoopptak av samtaler mellom islandske lingvister som deltok på NORMS-workshopen på Færøyene i 2008. Den tredje typen er lydopptak fra to samtaler med yngre informanter, tatt opp ved Háskóli Íslands. Alle opptakene er ortografisk transkribert.

Norsk: Opptakene har foregått i regi av prosjektet Norwegian Dialect Syntax (NorDiaSyn), et samarbeidsprosjekt mellom Universitetet i Oslo, Norges teknisk-naturvitenskapelige universitet og Universitetet i Tromsø. Prosjektet fikk finansiering fra flere kilder, og den norske delen av korpuset har derfor blitt den klart største. NorDiaSyn-opptakene består av videoopptak med nesten 440 informanter fra 111 steder i Norge. Fra hvert målepunkt deltok fire informanter, to yngre og to eldre, to kvinner og to menn. Informantene ble både intervjuet og var med på en samtale med en annen informant. Noen av opptakene ble gjort i forbindelse med

NORMS-workshoper, den første på Senja allerede i 2006. Den norske delen inneholder også en del eldre lydopptak fra Norsk målførearkiv. Opptakene er fra de samme dialektområdene som de moderne NorDiaSyn-opptakene, slik at det skal være mulig å bruke korpuset til diakrone studier.

Alle opptakene er transkribert på to måter: Først ble opptakene transkribert talemålsnært, deretter ble transkripsjonene oversatt til ortografisk transkripsjon ved hjelp av en halvautomatisk dialekt-transliterator utviklet ved Tekstlaboratoriet.

Svensk: Fra Sverige var vi så heldige å få bruke lydopptak og transkripsjoner fra et tidligere prosjekt, Swedia 2000. Dette består av intervjuer med 109 informanter, flest eldre, men også noen yngre, fra 31 ulike steder i Sverige. Opptakene er ortografisk transkribert. Resten av det svenske materialet ble tatt opp under en NORMS-workshop i Älvdalen i 2007. Dette er videopptak av intervjuer og samtaler som senere er transkribert etter den älvdalske skriftspråksstandarden, godkjent av det älvdalske språkrådet Rådöjärum. Transkripsjonene er deretter halvautomatisk oversatt til svensk standardortografi med dialekttransliteratoren som er utviklet for prosjektet.

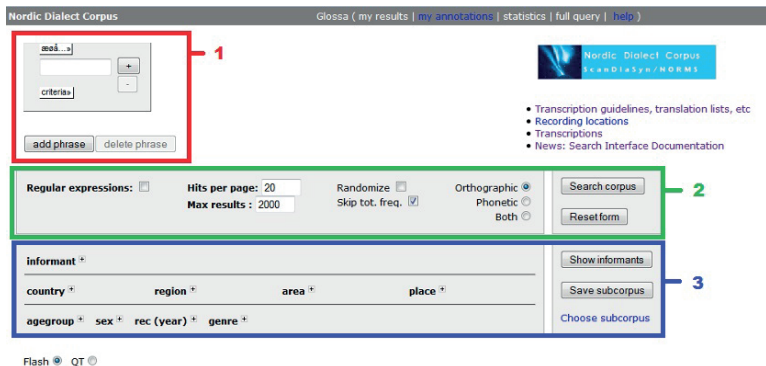
Søking i korpuset

Korpuset er teknisk utviklet ved Tekstlaboratoriet, UiO.⁴ Løsningene bygger dels på teknologi utviklet ved Tekstlaboratoriet for tidligere korpus, og er dels nyutviklet for *Nordisk dialektkorpus*. Korpuset er enkelt å bruke, men gir likevel mange avanserte muligheter. Søkesiden til *Nordisk dialektkorpus* består av tre hovedfelt, se bilde 1. Nedenfor skal vi illustrere bruken av korpuset med et enkelt eksempel: Vi ønsker å få en liste over alle treff på ordet *vi* i Valle i Setesdal.

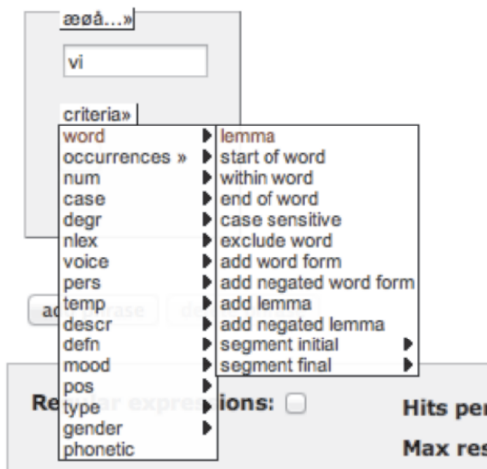
Felt 1 er det lingvistiske søkefeltet. Her kan man søke etter ett eller flere ord, lemmaer⁵, ordklasser eller utenomspråklige lyder som latter eller kremting. For vårt enkle formål må vi skrive ordet *vi* i søkefeltet øverst til venstre, bilde 2.

4. Hovedutvikler for Nordisk dialektkorpus har vært Joel Priestley, i samarbeid med Janne Bondi Johannessen, Anders Nøklestad og Kristin Hagen.

5. Bare norsk og dansk er lemmatisert, så lemma-kriteriet kan bare brukes for disse språkene.



Bilde 1: Søkevinduet til Nordisk dialektkorpus kan deles inn i tre felt. Under logoen kan man finne lenker til prosjektsiden, kart over opptakssteder og brukermanual. Det er også mulig å klikke seg inn og lese hele transkripsjoner uten lyd eller bilde.



Bilde 2: Vi ser ordet vi skrevet inn i søkeboksen, samtidig som nedtrekksmenyen også gir muligheten for å spesifisere videre.

I felt 2 kan man spesifisere hvordan søkeresultatet skal se ut. For transkripsjoner som har både talemålsnær og ortografisk transkripsjon, kan man for eksempel velge om man vil se bare én av transkripsjonene eller begge. Her velger vi begge, for å enkelt å kunne lese av variasjonen.

Felt 3 er spesifikasjonsfeltet for søking i metadataene i korpuset. Her kan søket begrenses til å gjelde noen utvalgte informanter eller informanter fra et spesifikt geografisk område. Man kan også bruke opptaksår, kjønn, aldersgruppe, intervju eller samtale som søkekriterium. Vi velger her Valle i Setesdal for vårt eksempel.

Resultatsiden for korpuset





Resultatene av et søk vil åpne seg i et nytt vindu, se bilde 3. Resultatene blir vist som en konkordans, med ett treff per linje, sammen med informantnavn og to eller tre klikkbare symbol: ett for informasjonsvindu, ett for videovisning og ett for lyd. Dersom man har spesifisert at man vil se begge transkripsjonene, blir den ortografiske transkripsjonen vist på linjen over den lydne.

CWB expression: "(((word="vi" %c))) "





Action :

Hits found: **163**





Results pages: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#)

    **valle_02uk** det er så det fungerer da kommer alle og så er det heim i tre uker og så sender **vi** :
de e så de fonggerer då kjem adde å så e de hæim i træi viko å så sende **mi** adde

translate


    **valle_02uk** ja jeg tycker det # jeg e ## jeg stortrives egentlig med # med de sauene **vi** har da
ja eg tikkje de # eg ee # eg stortrævast æigentleg me # me dæi saoin **mi** have då

translate

    **valle_02uk** ja # så det **vi** har ikke heilt klart å la oss rive oss helt i fra # fra byen så vi har fort
ja # så de **mi** hav ikkje heilt klart å la åss riv åkkå hæilt i frå # frå boien så mi ha !

Bilde 3: Eksempel på et søkeresultat der to transkripsjoner vises (her: ordet vi avgrenset til Valle i Setesdal). Øverst på siden kan man se hvor mange treff søket har gitt. Action-menyen gir ulike muligheter til å vise, bearbeide eller lagre søkeresultatene. Map viser resultatene i et kart.

Søket vårt ga også mulighet for å søke etter forskjellige bøyingsformer (både subjekts- og objektsformene av *vi* for eksempel), og resultatene vil da vise eksempler på *mi* (for standard *vi*), men også *åkkko* for *oss*. Dersom man klikker på et video- eller lydssymbol, får man opp en mediespiller, se bilde 4. Konteksten kan utvides.



tekstlab.uio.no/cgi-bin/glossa//query_dev.cgi#

valle_02uk:	* ja
valle_02uk:	så så da gjør han veldig mye av det derre ...
valle_01um:	* ja ja * (uninterpretable) * jo men
valle_02uk:	det er nå godt å holde seg noe i aktivitet (uninte
valle_01um:	* ja da så det trur
valle_02uk:	jeg han bare tykker er gama så # så hvis han g
valle_01um:	* ja * ja
valle_02uk:	men det er lamminga da den er jo travel ## så
valle_02uk:	vi så da
valle_01um:	* ja den er jo * det er (uninterpretable) * ja * ja
valle_01um:	og da er det fullt opp
valle_02uk:	da f- f- heile døgnet så det ...
valle_01um:	* ja ja
valle_01um:	hvor mange har du nå på ja ikke det var jo ikke
valle_01um:	eller hva ... ?

Bilde 4: Mediespilleren med utvidet søkeresultat til høyre. To unge fra Setesdal snakker.

Resultatene kan også vises på kart. Dersom transkripsjonen finnes i både lydner og ortografisk versjon, får man også muligheten til å se den geografiske utbredelsen av de lydnære variantene av søkeordet, slik bilde 5 på neste side viser. For å få litt glede av kartfunksjonen har vi denne gangen utvidet søket til hele Norden (og ikke bare Valle), samt tatt bort kravet til at *vi* skal være lemma. Variasjonen i forskjellige måter å uttrykke *vi* på, utnyttes på denne måten til å få fram automatiske isoglosser (geografiske skillelinjer for språkdrag).

Vi har nå vist hvordan et stort materiale kan bli lett tilgjengelig gjennom et korpus, slik *Nordisk dialektkorpus* er. Noen tall og fakta om korpuset samt et eksempel på bruksmåte håper vi har illustrert hvor nyttig dette korpuset kan være for språkforskere nesten uansett faglig interesse og forskningsfelt. Det mener vi også artiklene i denne boka illustrerer.



Bilde 5: Søk på vi: svart = oss, prikk = andre uttalevarianter av vi. Det kommer tydelig fram at det er et øst–vest-belte av dialekter hvor man sier oss der andre dialekter bruker en variant av vi i subjektform.

Takk

Denne boka kunne ikke blitt til om det ikke var for alle informan-
tene i de fem landene som har stilt seg til rådighet for intervjuer og

samtaler, slik at korpuset kunne bli en realitet. Disse skal ha vår største takk.

Det har vært viktig for oss at kvaliteten på de faglige artiklene har blitt best mulig, slik det er vanlig i vitenskapen. Derfor har vi bedt fagfolk om å vurdere artiklene kritisk, slik at de kunne bli enda bedre. Vi er svært takknemlige overfor disse, og nevner dem her: Ragnhild Anderson (Universitetet i Bergen), Maia Andreasson (Göteborgs universitet), Sturla Berg-Olsen (Universitetet i Oslo), Helge Dyvik (Universitetet i Bergen), Kristin Melum Eide (Norges teknisk-naturvitenskapelige universitet), Atle Grønn (Universitetet i Oslo), Marit Julien (Lunds universitet), Helge Lødrup (Universitetet i Oslo), Madeleine Halmøy (Universitetet i Tromsø – Norges arktiske universitet), Toril Opsahl (Universitetet i Oslo), Arne Torp (Universitetet i Oslo), Mai Tungseth (Universitetet i Tromsø – Norges arktiske universitet) og Mike Whelpton (Islands universitet).

Dessuten vil vi gjerne takke dem som har støttet arbeidet med dialektinnsamlingen og den tekniske korpusutviklingen økonomisk: Norges forskningsråd, Nordens forskningsråd, NOS-HS, SpareBank 1 Nord-Norge, Universitetet i Oslo, Norsk Ordbok 2014 og dessuten Directorate of Labour (Vinnumálastofnun) og Institute of Linguistics (Málvísindastofnun Háskóla Íslands), Forskningsrådet for Kultur og Kommunikasjon og Vetenskapsrådet.

Til slutt takker vi Institutt for lingvistiske og nordiske studier, Universitetet i Oslo, for støtte til bokutgivelsen.

Referanser

Johannessen, Janne Bondi, Lars Nygaard, Joel Priestley og Anders Nøklestad 2008: *Glossa: a Multilingual, Multimodal, Configurable User Interface. Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*. Paris: European Language Resources Association (ELRA).

Johannessen, Janne Bondi, Joel Priestley, Kristin Hagen, Tor Anders Åfarli, og Øystein Alexander Vangsnes 2009: The Nordic Dialect Corpus – an Advanced Research Tool. Jokinen, Kristiina and Eckhard Bick (eds.): *Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009. NEALT Proceedings Series*, Vol. 4 (2009), 73–80. Elektronisk versjon: Tartu University Library (Estland) <http://hdl.handle.net/10062/9206>.

Lindstad, Arne Martinus, Anders Nøklestad, Janne Bondi Johannessen og Øystein Alexander Vangsnes 2009: The Nordic Dialect Database: Mapping Microsyntactic Variation in the Scandinavian Languages. Jokinen, Kristiina and Eckhard Bick (eds.): *Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009. NEALT Proceedings Series; Volum 4*, 283–286. Elektronisk versjon: Tartu University Library (Estland) <http://hdl.handle.net/10062/9206>.

Korpussøkesystemet Glossa:

<http://www.hf.uio.no/iln/tjenester/kunnskap/sprak/glossa/>

Nordisk dialektkorpus og Nordisk syntaksdatabase:

<http://www.tekstlab.uio.no/nota/scandiasyn/>

Tekstlaboratoriet:

<http://www.hf.uio.no/iln/om/organisasjon/tekstlab/>

Janne Bondi Johannessen

Tekstlaboratoriet og MultiLing
Institutt for lingvistiske og nordiske studier
Universitetet i Oslo
P.b. 1102 Blindern
0317 Oslo
jannebj@iln.uio.no

Kristin Hagen

Tekstlaboratoriet
Institutt for lingvistiske og nordiske studier
Universitetet i Oslo
P.b. 1102 Blindern
0317 Oslo
kristin.hagen@iln.uio.no