



# LIA-korpuset som ressurs i revisjonen av tre ordbøker

Bente Selback og Terje Svardal

I denne artikkelen ser vi nærmare på kva ressurs eit talemålskorpus som LIA norsk kan vere i arbeidet med revisjonen av tre ordbøker: standard-ordbøkene *Bokmålsordboka* og *Nynorskordboka* og dokumentasjonsordboka *Norsk Ordbok*. I revisjonen arbeider ein både med å finne fram til nye lemma som bør takast inn i ordbøkene, og ein vurderer dei ordbokartiklane som er med frå før. Redaktørane må mellom anna ta stilling til om eit ord er så sentralt at det bør vere med vidare, om definisjonane er i tråd med gjengs språkbruk, kva bruksdøme som vil fungere best, og kva som er dei viktigaste kollokasjonane. Det er naudsynt å ha gode kjelder, og i artikkelen vurderer vi kva ein kan og ikkje kan bruke LIA-korpuset til i arbeidet.

Stikkord: leksikografi, ordbokrevisjon, skriftspråk, talemål, korpus

## 1 Innleiing

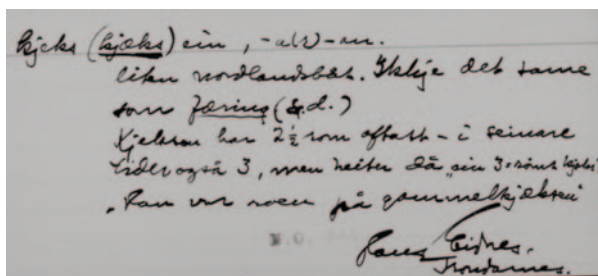
Kvardagen til ein leksikograf består av ei endelaus rekkje med spørsmål: Er dette lemmaet frekvent nok til å få ein plass i ordboka? Korleis kan ein lage ein god definisjon? Kva døme skal ein velje for illustrere bruken av ordet på ein forståeleg og representativ måte? Kva kollokasjonar er dei vanlegaste? Er ordet del av eitt eller fleire faste uttrykk? For å finne svar på desse og mange andre spørsmål treng ein gode og varierte kjelder som kan vise korleis orda blir brukte i både skrift og tale. Kva kjelder som eignar seg best, kjem sjølvstakt an på kva spørsmål ein stiller, og kva ordbok ein arbeider med. I denne artikkelen vil vi ta utgangspunkt i revisjonen som er i gang av standardordbøkene *Bokmålsordboka* og *Nynorskordboka* og av dokumentasjonsordboka *Norsk Ordbok*. Kan eit talemålskorpus som LIA norsk brukast i dette arbeidet, og kva kan det i så fall brukast til? Sidan det er snakk om to ulike ordboktypar, skal vi sjå på standardordbøkene og *Norsk Ordbok* kvar for seg og vurdere kva nytte ein kan ha av LIA i revisjonen av dei eksisterande artiklane og for å finne fram til nytt ordtilfang. Før vi ser nærmare på desse spørsmåla, vil vi kort vise kva overgangen til digitale materialsamlingar har hatt å seie for ordbokarbeidet, og presentere dei tre ordbøkene og revisjonen.

## 2 Leksikografi før og no – frå setelarkiv til elektroniske korpus

Mykje har endra seg sidan arbeidet med *Norsk Ordbok* og seinare standardordbøkene tok til. Tidlegare måtte leksikografane gå gjennom ordsetlane i arkivet for å finne svar på ulike spørsmål. Var ein til dømes ute etter kva ulike tydingar ordet *kjeks* kunne ha, måtte ein hente fram den aktuelle arkivskuffa, lese gjennom dei ofte handskrivne setlane og sortere dei i bunkar etter tydingane. I dag har ein god hjelp av elektroniske korpus som inneheld store mengder tekst, og der ein har avanserte søkjemoglegheiter som ofte gjer det råd å raskt finne dei opplysningane ein leiter etter.

## 2.1 Setelarkivet

Setelarkivet til *Norsk Ordbok* (NO-arkivet) omfattar kring 3,4 millionar setlar med kring 500 000 unike lemma (Vikør 2018, 32).<sup>1</sup> Arkivet vart i hovudsak bygd opp frå 1930 og var i jamn vekst til midt på 90-talet. Frå 2002 har det vorte produsert elektroniske setlar. Materialet i NO-arkivet er samla inn dels av ordbokredaksjonen og dels av frivillige medarbeidarar. Dermed er det nokså ueinsarta. Om lag 600 frivillige medarbeidarar landet rundt bidrog, og dei har levert alt frå éin til kring 65 000 setlar. Variasjonen i bakgrunn og interesser har vore relativt stor, noko materialet ber preg av. Det varierer desutan kor mykje tilfang det er frå dei ulike målføra. Eit område som enno er lite dekt, er kyststroka av Agder.



Figur 1: Ein av dei 92 arkivsetlane med ordet *kjeks*

Det finst to hovudgrupper av setlar, dei med litterære ekserpt og dei med bruksdøme frå målføra. Om lag ein halv million av setlane inneheld opplysningar frå målføra om ordformer, bøyning, tydingar og døme på bruk (Vikør 2018, 32). Arkivet finst både som tradisjonelle arkivsetlar, som er skanna og gjort tilgjengelege av Dokumentasjonsprosjektet på 1990-talet (Kristiansen og Ore 1998), og som nyare elektroniske setlar.

<sup>1</sup> Setelarkivet til *Norsk Ordbok*: <http://usd.uib.no/perl/search/search.cgi?tabid=436&appid=8>

## 2.2 Elektroniske skriftspråkskorpus

Sjølv om ordsetlane framleis er i bruk, nyttar vi i dag primært elektroniske korpus i det leksikografiske arbeidet. Det finst mange ulike skriftspråkskorpus som femner om ulike typar tekst. Særleg korpus som inneheld tekstar som har vore gjennom ein redaksjonell prosess, er interessante for oss. Aller mest bruker vi korpushandteringssystemet Korpuskel (Meurer 2012), som er vidareutvikla for ordbokrevisjonen gjennom Korpuskel-Lex.<sup>2</sup> Korpuskel-Lex har avanserte søkjemoglegheiter og gjev oss høve til å søkje på tvers av ei samling ulike korpus, mellom anna Norsk aviskorpus og Nynorskkorpuset (Lyse 2020). Til saman inneheld skriftspråkskorpusa i Korpuskel-Lex om lag 2,7 milliardar ord. Det gjev oss betre moglegheiter enn tidlegare til å finne fram til den representative språkbruken. Samstundes kan materialet bli overveldande. Eit trunkert søk på ord som startar med *kjeks*, gjev til dømes 7783 treff, noko som gjer det vanskeleg å få oversikt. Heldigvis er det mogleg å snevre inn søket slik at ein får eit meir handterleg materiale.

nnk	↑ sett fram saft og kjeks, og eg har smakt på	kjeks	, mest for å vera høflig. – Sjå her! Her er me	2009
nnk	verdas dyraste kjeks gav meg, den påskribla	kjeks	frå Sir Ernest Shackletons mislykka ekspedis	2008
nnk	rødde istappar, og så midt opp i dette denne	kjeks	, så komplett absurd, og no kjem alle dei stu	2008
nnk	, fersk kokfisk, meinte eg. Og vesleguten tek	kjeks	, eller ka veslebaaten no heiter, og dreg utp:	1911
nnk	det er mange ting å finne rom åt i den vesle	kjeks	eller spissbåten. Båtene må lastast rett, så h	2005
nnk	med kaffikrem, og i valnott- og kokoskaker.	Kjeks	er produserte av eit japansk selskap, medan	2008

Figur 2: Utdrag frå konkordans i Korpuskel-Lex som viser treff på *kjeks*.\* i Nynorskkorpuset

## 2.3 Elektroniske talemålskorpus

Dei fleste og største korpusa inneheld skriftspråk, men det finst òg fleire tilgjengelege korpus med transkribert talemålsmateriale. Her vil vi berre nemne to av dei vi har brukt mest i revisjonsarbeidet, Nordisk dialektkorpus og Talebanken.<sup>3</sup> Dei eldre delane av Nordisk dialektkorpus er flytta over i LIA-korpuset, men den norske delen

<sup>2</sup> Korpuskel-Lex: <https://clarino.uib.no/lex/>

<sup>3</sup> Nordisk dialektkorpus: <https://tekstlab.uio.no/glossa2/ndc2>

Talebanken: <https://clarino.uib.no/korpuskel/corpus-list?collection=Talebanken>

består framleis av opptak frå 111 stader rundt i landet. Opptaka er gjorde i perioden 1998–2015, og dei er transkriberte i både ein tale-målsnær variant og til bokmål. Korpuset er på om lag 2 millionar ord. Talebanken inneheld dei tre korpusa *Dialektendring*, *Industri-stad* og *Talesøk*. Han har både eldre og nyare opptak, primært frå Vestlandet. Opptaka er transkriberte til nynorsk. Talebanken inneheld om lag 7,6 millionar ord.

Sjølv om det alt finst fleire talemålskorpus, har vi ikkje like gode korpusressursar for tale som for skrift. Eit talemålskorpus kan seie oss kva ord og former som er brukte i målføra rundt om i landet, noko som er nyttig for *Norsk Ordbok*. Men det kan òg fortelje oss noko generelt om kva som kjenneteiknar talemål i motsetnad til skriftspråk, og slik vere eit godt supplement til skriftspråkskorpusa. Både LIA-korpuset, Talebanken og Nordisk dialektkorpus inneheld munnleg språkbruk i form av dialogar. Vi kan tenkje oss at det vil vere ei god kjelde til ordtilfang som er typisk for samtale, som preposisjonar, interjeksjonar, pragmatiske partiklar og så vidare, og at det difor særleg kan vere til hjelp i revisjonen av denne typen ord.

### 3 Ordbøkene og revisjonen

I 2016 tok Universitetet i Bergen over ansvaret for Språksamlingane frå Universitetet i Oslo. Samlingane inneheld store mengder fysisk og elektronisk materiale innanfor fagfelt som namnegranskning, norrønt, dialektologi og leksikografi. Saman med det leksikografiske materialet fekk ein ansvaret for *Norsk Ordbok*, *Bokmålsordboka* og *Nynorskordboka*. Ordbøkene er ikkje ferdige ein gong for alle. Verda og språket endrar seg som kjent, og då er det nødvendig å revidere ordbøkene òg. Difor er det i gang eit revisjonsarbeid for alle dei tre verka.

*3.1 Standardordbøkene Bokmålsordboka og Nynorskordboka*  
*Bokmålsordboka* og *Nynorskordboka* vart tidlegare kalla *handord-bøkene*. Dette omgrepet viser til at dei var meinte å vere eittbands ordbøker i eit nokolunde hendig format. I dag er ordbøkene elekt-

roniske, og dermed er ikkje plassomsyn like viktig som i ei trykt bok, men dei to ordbøkene skal framleis vere to mellomstore allmennordbøker for dei moderne skriftspråka bokmål og nynorsk. Dette er styrande for omfanget og lemmautvalet.

Standardordbøkene er mykje brukte. I 2020 vart det gjort nærmare 58 millionar søk i dei elektroniske ordbøkene, og fleirtalet av brukarane vel standardsøket som gjev søkjeresultat i begge ordbøkene samstundes (Ordbokstatistikk 2020). Brukarane av ordbøkene inkluderer alt frå profesjonelle skribentar til skuleelevar og personar som prøver å lære seg norsk som andrespråk. Den samansette brukargruppa er noko vi må ha i tankane når vi utformar ordbøkene.

Historia til standardordbøkene tok til i 1966. Då kom innstillinga frå Vogt-komiteen, og eitt av framlegga deira var å opprette eit norsk leksikografisk institutt. Dei skreiv vidare at «Det som i dag trenges kanskje mere enn noe annet, er gode, praktiske ettbinds ordbøker til overkommelig pris for bokmål og nynorsk, eventuelt for begge språkformer under ett» (Innstilling 1966, 13).

Norsk leksikografisk institutt vart oppretta i 1972 og lagt til Universitetet i Oslo. Arbeidet med dei to ordbøkene tok til i 1974 i samarbeid med Norsk språkråd (Vikør og Wetås 2016, 23–24). Arbeidet var organisert slik at det var to redaksjonar, ein for kvar målform, men dei samarbeidde om opplegget og delte alfabetet mellom seg slik at dei starta på kvar sine delar og sidan bytte manus. Denne framgangsmåten skulle sikre at dei to ordbøkene vart mest mogleg like kvarandre, men då førsteutgåvene var klare i 1986, viste det seg at ordbøkene likevel ikkje hadde vorte så like som ein først hadde tenkt. Dette hadde samanheng med ulikt kjeldegrunnlag og ideologiske og historiske ulikskapar mellom målformene. Mellom anna var det viktig for nynorskredaksjonen å få med ein større del talemålstilfang (Worren 1998). *Nynorskordboka* med om lag 90 000 oppslagsord vart difor ein del større enn *Bokmålsordboka* med om lag 65 000 oppslagsord.

Begge ordbøkene er komne i fleire nye utgåver sidan den gongen. I 1994 vart dei tilgjengelege i elektroniske utgåver på nettet (ordbok.uib.no). Nettutgåva av *Nynorskordboka* vart oppdatert i 2012 og *Bokmålsordboka* i 2016. Då fekk dei ny utsjånad og funksjonali-

tet, med ordklasse og leddeling på alle oppslagsord, søkbare fleirordsuttrykk, hyperlenking osv. I 2017 kom dei som appen *Ordbøkene*. Rettskriving og bøyingar er oppdaterte i samband med nyare rettskrivingsvedtak, og ein del nyord er lagde til, men ein har ikkje gått systematisk gjennom lemmautvalet og innhaldet i artiklane sidan førsteutgåvene (Hovdenak 2013).

Arbeidet med revisjonen av begge ordbøkene vart sett i gang i 2018. Prosjektet er eit samarbeid mellom Universitetet i Bergen og Språkrådet, og det har fått økonomisk støtte frå Kulturdepartementet. Arbeidet er organisert litt annleis enn tidlegare. Denne gongen er det berre éin redaksjon som arbeider med begge ordbøkene parallelt. Tre viktige oppgåver er å

- redusere skilnader mellom dei to ordbøkene som ikkje har grunnlag i faktiske målformskilnader
- syte for at definisjonar og bruksdøme i eksisterande artiklar er i tråd med språkbruken i dag
- ta inn nye oppslagsord

### 3.2 Dokumentasjonsordboka Norsk Ordbok

*Norsk Ordbok – Ordbok over det norske folkemålet og det nynorske skriftmålet* vart til i tida mellom 1930 og 2015. Dette er eit vitenskapleg ordbokverk som omfattar 12 band og inneheld kring 330 000 artiklar. Formålet med *Norsk Ordbok* (NO) er å gje ei uttømmende vitenskapleg framstilling av ordtilfanget i dei norske dialektane frå 1600 til i dag og i det nynorske skriftmålet (Vikør 2018, 10). I artiklane finn ein ordtydingar, etymologiske opplysningar, bruksdøme frå litteraturen og målføra, dialektformer og grammatiske opplysningar. Definisjonane byggjer på grundig semantisk analyse av eit datamateriale og kan ha kjeldefeste døme frå litteraturen og målføra. Idealet er at verket skal skildre alle aspekt av språket, dekkje alle delar av samfunnet og alle sjangrar. Verket vender seg til vaksne språkbrukarar med norsk som morsmål, vanleg skulegang og sterk interesse for norsk språk. Særlege målgrupper er lærarar, vitkapsfolk og lekfolk med interesse for målføre, nynorsk litteratur og språktradisjon (Vikør 2018, 10).

Verket vart fullført i 2015, men artiklane *a* til *håvøyll* er ikkje digitalt tilgjengelege. Prosjektet NO-AH (Revisjon og digitalisering av *Norsk Ordbok a–h*) tok til i 2019. Det er eit samarbeid mellom Universitetet i Bergen og Høgskulen i Volda. Målet er å gjere heile verket digitalt tilgjengeleg på nettet. Prosjektet omfattar oppdatering av redaksjonelt innhald og tekniske løysingar. Frå første hefte av band 1 vart utgjeve i 1950, har det skjedd endringar i språket og samfunnet. Vi har utvikling til dømes på det tekniske området, der eit felt som data- og kommunikasjonsteknikk har endra seg sterkt og ordtilfanget med dette.

I tillegg til innhaldet er det behov for endring av databaseløysing og redigeringsgrensesnitt. Dei tekniske spørsmåla går vi ikkje inn på i denne artikkelen.

## 4 LIA-korpuset og revisjonen av standardordbøkene

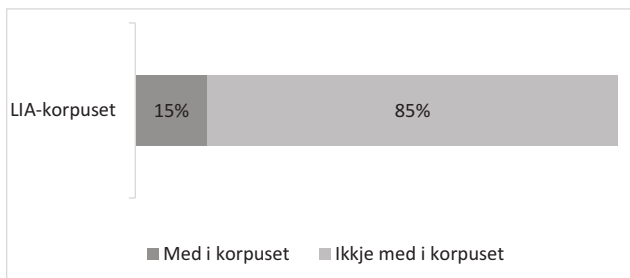
Ei ordbok er eit utval av eit prinsipielt uendeleg ordtilfang, og kva type ordbok det er snakk om, er med på å avgjere ordutvalet. Som nemnt ovanfor er *Bokmålsordboka* og *Nynorskordboka* middels store allmennordbøker. I teorien er dei normative på rettskriving og bøyning, men ikkje i lemmautval, men i praksis blir lemmautvalet ofte oppfatta normativt av brukarane. Dersom eit ord ikkje står i ordbøkene, vil mange oppfatte det som at ordet dermed er «forbode» å bruke, noko mange av spørsmåla redaksjonen får, viser (Rauset 2019). Vi autoriserer ikkje ordtilfanget, men vel ut dei lemmaa vi meiner det er vesentleg å beskrive i ordbøkene. For at utvalet skal bli så representativt som råd, treng vi gode kjelder, og vi skal sjå nærare på om LIA-korpuset kan vere ei slik kjelde.

### 4.1 Revisjon av eksisterande artiklar i standardordbøkene

I revisjonen av standardordbøkene er arbeidet organisert på den måten at kvar redaktør får tildelt ein alfabetbolck som han eller ho så reviderer i begge ordbøkene. Ein vanleg arbeidsmåte er å gå systematisk gjennom ordbokartiklane i bolcken og søkje opp orda i korpus for å mellom anna sjekke kor utbreidd bruken av dei er, om definisjonen



ser ut til å stemme, og om døma er gode. For å undersøkje om LIA-korpuset eignar seg til denne arbeidsmåten, har vi gjort ein liten test av dei 300 første orda i ein tilfeldig alfabetbolk. Bolken startar med ordet *aviarium* ‘fuglebur’ og sluttar med *avsanne*. Vi har søkt systematisk i korpuset på alle oppslagsorda i bolken for å sjå kor mykje informasjon ein kan finne. Figur 3 viser resultatet av testen.



Figur 3: Test av ein ordbokbolk i LIA-korpuset (N = 300)

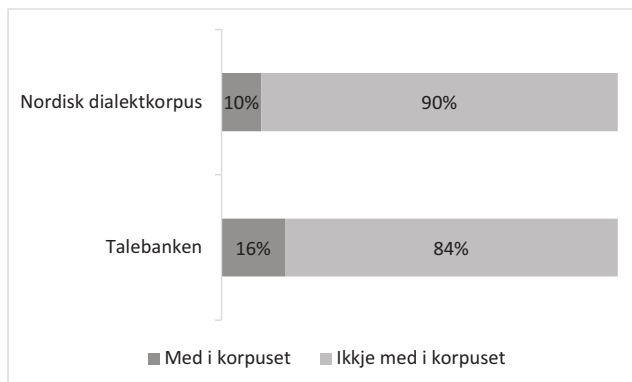
Av dei 300 undersøkte orda var det 45 som gav tilslag i LIA-korpuset, altså berre 15 prosent, og over halvparten av dei orda vi fekk tilslag på, hadde berre eitt eller to belegg. Det betyr at det for dei fleste eksisterande ordbokartiklane ikkje er så mykje informasjon å hente i dette korpuset, med unntak av at vi kan slå fast at dei fleste oppslagsorda ikkje er frekvente i talemål.

Vi har testa den same bokstavbolken i Talebanken og den norske delen av Nordisk dialektkorpus. Heller ikkje i desse talemålskorpusa får vi mange treff (sjå figur 4 neste side). Talebanken gjev om lag like mange treff som LIA-korpuset med treff på 16 prosent av oppslagsorda. I Nordisk dialektkorpus får vi ikkje treff på meir enn kvart tiande oppslagsord.

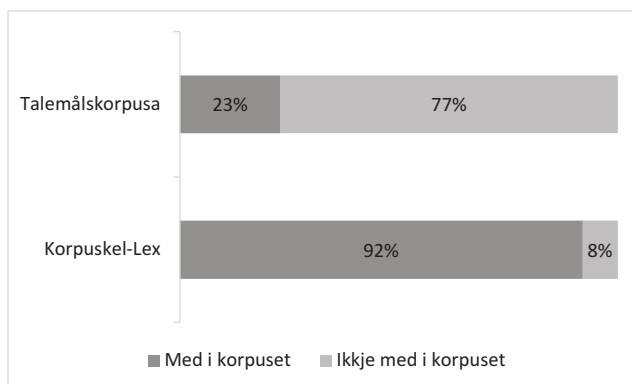
Om vi slår saman treffa for alle dei tre talemålskorpusa, hjelper det litt (figur 5 neste side), men framleis får vi treff på mindre enn kvart fjerde ordbokoppslag. Til samanlikning gjev søk i skriftspråkskorpusa i Korpuskel-Lex ein treffprosent på over 90.

Det er kan hende ikkje så veldig overraskande at vi får få treff når vi testar eit skriftbunde ordtilfang mot ein talemålsressurs på denne

måten. Éi mogleg årsak er storleiken på korpusa. Dei tre undersøkte talemålskorpusa er til saman på om lag 13 millionar ord. Til samanlikning er Nynorskorpuset på over 100 millionar ord og dei samla skriftspråkskorpusa i Korpuskel-Lex på om lag 2,7 milliardar ord. Ei slik skeivfordeling må vi nok berre leve med sidan eit talemålskorpus krev mykje meir arbeid å byggje opp.



Figur 4: Test av ein ordbokbolk i Nordisk dialektkorpus og Talebanken (N = 300)



Figur 5: Test av ein ordbokbolk i tre talemålskorpus (LIA, Talebanken og Nordisk dialektkorpus) og i skriftspråkskorpusa i Korpuskel-Lex (N = 300)

Storleiken på korpusa er likevel ikkje heile forklaringa på dei manglande treffa. Viss vi ser nærmare på kva ord den undersøkte bolken inneheld, er vel 70 prosent substantiv, om lag 20 prosent verb og resten adjektiv og adverb. Dette vil sjølvstundt variere noko alt etter kva alfabetstrekk det er snakk om, og andre strekk vil ha fleire funksjonsord, men det er ingen tvil om at det er innhaldsorda som dominerer ordbøkene.

Ordbøkene har med mange ord som er lite vanlege i daglegspråket. Dei seks første orda i den undersøkte bolken er *aviarium*, *aviatkar*, *aviatikk*, *aviatisk*, *avig* og *aving*, og det må seiast å vere ord dei færreste av oss bruker jamleg, men som det likevel kan vere fint å finne i ordboka dersom ein kjem over dei og lurar på kva dei tyder.

Ordbøkene inneheld dessutan veldig mange samansetjingar, noko som kan vere nyttig, særleg der samansetjingane er lite gjennomsluttede, eller der det kan vere vanskeleg å vite om orda skal ha samansetjingsfuge. Heiter det til dømes *ordbokredaktør* eller *ordboksredaktør*? Ein del av dei samansette orda som er med i dag, er nokså tilfeldige eller utdaterte. Den undersøkte alfabetbolken inneheld ord som *avisklipparkiv*, *avkledningsscene*, *avkomsgransking*, *avlatskremmar*, *avlusingsanstalt*, *avmagringsmiddel* og *avrekningskontor*. Det er relativt lite sannsynleg at vi skal komme over samansetjingar som desse i eit talemålskorpus.

Dei manglande treffa kan òg ha å gjere med kva materiale som er lagt inn i talemålskorpusa. Alle dei undersøkte korpusa er baserte på opptak av intervju med informantar, og slike intervju er ein spesiell sjanger. Det er visse emne ein ofte snakkar om, mens andre emne sjeldan blir nemnde, og det vil sjølvstundt påverke kva ord vi finn i korpuset. Hadde korpusa vore sette saman av radioopptak eller stortingsforhandlingar, ville vi funne eit anna ordtilfang. Dersom vi sorterar orda i alfabetbolken etter domene, vil vi til dømes sjå at mange ord som kan seiast å høyre til livet på garden, er å finne i korpusa (*avkom*, *avl*, *avle*, *avling*, *avlingsmetode*, *avlstdyr*, *avløysar*), mens det derimot ikkje er nokon som snakkar om dei mange språkvitskaplege termene som har fått plass i ordbøkene (*avleiing*, *avleiingsending*, *avleiingsforstaving*, *avleiingsprefiks*, *avleiingssuffiks*, *avlyd*, *avlydshøve*, *avlydssteg*, *avlydsveksel*, *avløysarord*).

Testen viser at ein framgangsmåte med automatiske søk i talemålskorpus på alle eksisterande oppslagsord i ordboka er lite tenleg. LIA-korpuset kan rimeleg nok ikkje hjelpe oss med alle typar ord, men det kan likevel vere nyttig å gjere meir målretta søk på ord der talemålskorpus har kvalitetar skriftspråkskorpusa ikkje har.

Standardordbøkene, og då særleg *Nynorskordboka*, inneheld ein god del tradisjonsord og talemålsord. Ein del av desse har vi få skriftlege belegg på, og det kan vere vanskeleg å vurdere om dei fortener ein plass i ei allmennordbok eller ikkje. Er ordet framleis i bruk i talemåla rundt om i landet, eller er det eit gammalt ord som ikkje er i bruk? Skal det bli ståande berre i *Nynorskordboka*, eller er ordet så vanleg at det bør opprettast ein tilsvarande artikkel for bokmål? Her kan tilgang til talemålskorpus hjelpe oss nærmare eit svar.

Eit døme er artikkelen *flautstokk*, som frå før berre er med i *Nynorskordboka*. Ein flautstokk er ein slede, og det er eit typisk døme på den typen ord det finst fleire av i *Nynorskordboka* enn i *Bokmålsordboka*. Men er dette eit ord som er vanleg i bruk? Dersom vi søker i Korpuskel-Lex, får vi ingen treff på dette ordet. Eit søk i Bokhylla til Nasjonalbiblioteket gjev ingen treff i aviser og vel 50 boktreff, men dei fleste av dei er treff i diverse ordbøker og ordsamlingar i tillegg til eit par bøker om skogsdrift i gamledagar. Vi kjem altså ikkje så mykje nærmare ei avklaring av om dette er eit ord som høyrer heime i ei moderne allmennordbok. Dersom vi søker opp ordet i LIA-korpuset, får vi derimot heile ti treff fordelt på tre informantar frå Brandbu, Nannestad og Tolga. Det tyder på at dette har vore eit vanleg omgrep i talemålet fleire stader. For to av informantane kjenner vi fødselsåret (1893 og 1903). For den tredje manglar denne informasjonen, men vi veit i alle fall at han var ein vaksen mann i 1968 då opptaket vart gjort, så vi kan tippe at han er fødd om lag på same tida som dei andre. Dei to andre opptaka er frå 1970 og 1973. LIA-korpuset fortel oss altså at ordet har vore brukt i talemålet til eldre folk på byrjinga av 70-talet, men om det framleis er i bruk i dag om lag 50 år seinare, kan ikkje korpuset fortelje oss. Til det er det meste av materialet i eldste laget.

Ei anna gruppe ord der eit talemålskorpus kan vere særleg nyttig, er i revisjonen av funksjonsorda. Desse er ein sentral del av ordtil-

fanget sjølv om dei ikkje utgjer ein stor prosent av ordbøkene. Ein del av dei har alt fått etter måten store artiklar, men dei kan justerast og supplerast ved hjelp av talemålsmateriale. Dessverre har ikkje prosjektet ressursar til å gjere djuptgåande analysar av store mengder nytt materiale, til det er det alt for mange artiklar som skal gjennomgåast på kort tid. Her må vi i staden stø oss på andre sitt arbeid. I dei seinare åra er det komme mange studiar som bruker data frå talemålskorpus, til dømes Søfteland og Borthen (2018) som har undersøkt den pragmatiske partikkelen *sjø* i Nordisk dialektkorpus.

#### 4.2 Nytt tilfang i standardordbøkene

I revisjonen skal vi ikkje berre gå gjennom dei artiklane som alt er med i ordbøkene, vi skal òg leggje til nye artiklar. Kan LIA-korpuset hjelpe oss med det? Når det gjeld *Nynorskordboka*, har transkribørane i LIA alt gjort ein del av jobben for oss. Dei talemålsnære transkripsjonane er translittererte til nynorsk, og materialet er lagt til rette slik at alle ord som ikkje er funne i *Nynorskordboka*, er merkte med ein x. Det betyr at det er mogleg for oss å søkje opp alle dei merkte orda og slik få ei liste over moglege kandidatar til ordbøkene. At eit ord er med på x-lista, vil ikkje automatisk seie at det skal takast inn i standardordbøkene, men det er eit godt utgangspunkt for å sjå nærare på ordet.

Dei to mest frekvente lemmaa på x-lista er *derre* (3141 treff) og *herre* (1434 treff). (1) viser to døme frå korpuset på korleis dei er brukte.

- (1) a) ja # tidig korleis dei liksom får dei derre # dei derre dokkene der til å # bevege seg sånn og (aasane\_uib\_0202)  
 b) ned # så køyrer du ned den herre bakken som heiter Storgjerdsbakken (alta\_uit\_0103)

Ruth Vatvedt Fjeld (2008, 17–18) etterlyste ordet *derre* i *Bokmålsordboka* etter ein analyse av NoTa-materialet. Sidan den gongen har det fått artikkel i *Bokmålsordboka*, men *derre* og *herre* manglar framleis i *Nynorskordboka*. Sjølv om mange av oss nok har lært at dette er eit talemålstrekk som bør lukast ut i skriftlege tekstar, gjev LIA-

korpuset oss eit hint om at vi bør undersøkje om orda likevel er i bruk i skrift. Eit søk i Korpuskel-Lex viser at begge orda er ein god del brukte, særleg i tekstar med eit munnleg eller uhøgtidleg preg. Før dei eventuelt kan komme med i ordboka, må Språkrådet ta stilling til om dei framleis skal vere unormerte i nynorsk (og altså ikkje takast inn i ordboka), om dei skal vere valfrie variantar av *derne* og *herne*, eller om dei skal reknast som separate lemma og få eigne artiklar.

Om vi elles ser på lista over x-orda, er det interessant å registrere at ho inneheld mange av dei orda som tradisjonelt har vore haldne utanfor nynorsken, mellom anna ei rekkje av dei såkalla anbeheitelse-orda (*antal*, *begynnelse*, *bebyggelse* osv.) og ord som *selyfølgelig*, *spise*, *kun* og *forrige*. For ord av tysk og dansk opphav er det Språkrådet som vedtek om dei kan takast inn i *Nynorskordboka* (Språkrådet 2015). Om vi ser vekk frå desse meir problematiske orda, kan x-lista hjelpe oss med å finne lakunar i ordbøkene. Av og til hender det at ord som er heilt vanlege og uproblematiske i språket, manglar. På x-lista finn vi til dømes *fjortis* og *tyggis*, og dei hadde ikkje artiklar i ordbøkene før revisjonen, men fortener nok ein plass i dei reviderte utgåvene.

## 5 LIA-korpuset og revisjonen av *Norsk Ordbok*

I prosjektet NO-AH skal ein leggje til rette for publisering av ei fullstendig utgåve av *Norsk Ordbok* på nettet. I arbeidet med revisjon og tilrettelegging for nettpublisering ser vi særleg to område der LIA-materialet kan brukast. Artiklar som allereie er trykte og finst i den digitale ordbokdatabasen, kan få tillegg av opplysingar knytte til seg. Dessutan kan vi identifisere lemma som til no ikkje har vore dokumenterte i språksamlingane som ordboka byggjer på.

Synet på kva som er «akseptabelt» folkemål å dokumentere, har kanskje endra seg sidan innsamlinga tok til på 1930-talet. Vi ser til dømes at det folkelege *dass* ‘toalett’ har fått artikkel. Det same har *dasshus*, men elles har ingen samansetjingar fått eigne artiklar. Tid-

legare undersøkingar i talemålskorpus peikar i retning av at til dømes *dasspapir* er frekvent nok til å få ein artikkel (Fjeld 2008, 19–20).

### 5.1 Revisjon av eksisterande artiklar i Norsk Ordbok

For eksisterande artiklar i verket kan LIA-materialet gje oss opplysningar som stadfester det vi veit frå før, eller det kan gje oss tilleggsopplysningar. Som døme fører vi *kjeks* i tydinga ‘båt’ (sjå figur 6).

**I kjeks m** [Nhl1750 *kiex*, G og Lof1807 *kjæx*, Lyngen-1810-12 *kæx*, C, A<sub>2</sub>; målf òg genus f (Sandstad, Lyngen); trul. frå eng. *ketch* 'eit slags tomastra seglbåt'] liten nordlandsbåt med to årepar og to (og eit halvt) rom (mindre enn ein færing) (Nhl1750; G; Lof1807; Lyngen1810-12; C; A<sub>2</sub>; Sandstad, NN): *han hadde så nett eit lite segl til tjeksen sin* (EidnesTT 129).

Figur 6: Artikkelen *kjeks* i *Norsk Ordbok*

Basert på funn i LIA-korpuset (2) får vi ei opplysning om tydinga som stadfester definisjonen som alt står i *Norsk Ordbok*. *Kjeks* kan brukast om ein båt.

- (2) så tok eg då å d- tjækksa ## en tjækksbåt å å dønn lå- laga ee laga sånn ee jåa ti mek # sånn tarrejåa (lyngen\_uit\_0702)

I artikkelen *kjeks* (sjå figur 6) står NN for Nord-Noreg, og i den elektroniske basen har vi heimfestingslista «Sandstad, Nesna, Lurøy, Fauske, Lofoten, Vesterålen, Bø i Vesterålen, Sør-Troms, Trondenes, Senja A<sup>2</sup>, Lyngen, Talvik, Måsøy, Lebesby, Gamvik». A<sup>2</sup> etter ‘Senja’ refererer til Aasens *Norsk Ordbog*. I LIA-korpuset (3) finn vi eit døme på at ordet er brukt av ein informant frå Andenes. Vi kan dermed leggje til heimfestinga Andenes, som vi ikkje hadde frå før.

- (3) å då sa hann M2 att hann sku ittje få låv å ro meire # åleina på de # tjækks'n di skulle # ta tjækks'n ifrå hann (andenes\_uio\_0201)

Med belegget frå Lyngen (2) får vi stadfest tydinga, og dessutan gjev dømet ei opplysning om genus femininum som vi kan føre i tillegg til opplysninga frå Sandstad som står frå før i den trykte artikkelen.

### 5.2 *Nytt tilfang i Norsk Ordbok*

Dei eldre samlingane av målføre er omfattande, men er prega av at materialet er samla av frivillige medarbeidarar. Sjølv med rettleiing frå ordbokredaksjonen vil det vere ulikt kva den einskilde har sendt inn. Talemålskorpusa vil difor kunne gje oss nytt tilfang å skrive ordbokartiklar om. Nye artiklar kan bli laga om ord som ikkje før har vore beskrivne, eller materialet kan gje opphav til nye artiklar på grunn av endra vurdering av kva ordklasse orda er tilordna.

Her hjelper annoteringa oss i og med at vi som nemnt ovanfor kan søkje på ordformer merkte med x. Desse kan takast ut som liste, og ho kan samanliknast med lemmalista frå *Norsk Ordbok*. Der vil ein kunne fange opp ord som ikkje tidlegare har artikkel i verket.

Ein mindre presis framgangsmåte, men som kan gje resultat, er å søkje med utgangspunkt i redaktøren sin kjennskap til innhaldet i *Norsk Ordbok* og kva tema som i mindre grad er omtalte i dei eldre samlingane. Vi veit at tradisjonelle kvinnedomene er dårlegare dokumenterte enn dei tradisjonelle mannsdomena. Difor gjorde vi eit enkelt søk på alle ord som byrjar med *bakst*. Dette søket gav oss dessverre ikkje nytt tilfang om bakst og baking, men det gav ein uventa bonus i form av ordet *bakstovelem* ‘loft over stove’. Ettersom sistelekken i denne samansetjinga er eit oppslagsord i *Nynorskordboka*, er ho ikkje med på x-lista, jamfør transkripsjonsrettleiinga for LIA. *Bakstovelem* fann vi i eit opptak frå Eikefjord i Sunnfjord (4).

- (4)   å i dennj anndre ænnda der va et # romm so dei kallte bakståva  
# elle bakståeva # åg på dennj va dær åg læmme så dei kallte  
bakståeulemmen # bakståeulemmen (eikefjord\_uib\_0102)

Ordet er ikkje registrert i samlingane frå før av, og funnet kan vere grunnlaget for ein ny artikkel i samband med revisjonen av *Norsk Ordbok*.



LIA-korpuset kan dessutan føre til ei ny vurdering av kva ordklasse eit ord høyrer til. Korpuset har døme (5 a–c) på bruk av preposisjonen *ani*:

- (5) a) inngen så turrde å vær ani hann (bergen\_uib\_0402)  
 b) mått ikkje vær ani pærsoner så add lus (bergen\_uib\_0402)  
 c) å så bare sånn så visst enn gutt å enn jænnte i klass'n bare vi kamma ani veranndre vi tullar sånn vi sann (bergen\_uib\_5102)

**II an** adv [mlty *an*, sm o s gno *á*, nno *á* 'på'; i tyd 3 frá ty *an*, kan henda ett ital *a* 'til'].  
**1)** lbr, i samb *an i* bort i, innåt: *koma, ná, vera an i*. **2)** i sume faste vend (vanl i talemål, lbr i skr): *gå an* vera godtakande; la seg gjera; serl i vend *det går (ikkje) an* det går (ikkje), ein kan (ikkje) | *koma an på* stå på, skilja seg: *det kjem an på om eg får lov* (òg: det spørst, det er tvilsamt) | *kom an* kom og slåst | *kom an, din rædde hund* (H.Rytt.Sh IV,95); òg med avveikt tyd, kom (og ver med) | *slå an a* (i samb med *med* el *på*), kvard, (av erotiske grunnar) gjeva seg i snakk (med) og prøva få fylgje (med ein person av anna kjønn), b) mus., lbr, gjeva, slå tone el akkord på eit strengeinstrument, serl fyrste tonen el akkorden i eit musikkstykk, c) verta omtykt, fengja. **3)** forretn, (i bokføring) nytta føre namnet til kreditor el (no forald) føre postane på debetsida i konto el på rekning; mots *per* (*pr.*).

Figur 7: Artikkelen *an* i *Norsk Ordbok*, band 1

**ani** prep i berøring med; borte i (Arendal, Brg); jf  
*II an 1: ein måtte ikkje vere ani personar som hadde lus* (Brg).

Figur 8: Forslag til ny artikkel for *ani* i *Norsk Ordbok*

Denne preposisjonen er registrert i samlingane frå før, men i den trykte utgåva av *Norsk Ordbok* er *ani* omtalt under artikkelen som omhandlar adverbet *an* (sjå figur 7, linje 3–4). Ein kan argumentere for å strukturere om informasjonen på den måten at preposisjonen *ani* får eigen artikkel. Resultatet av vurderinga så langt er eit utkast til artikkel i den digitale ordbokbasen (figur 8).

## 6 Oppsummering og konklusjon

Eit talemålskorpus som LIA er utan tvil eit nyttig verktøy for ein leksikograf. Kor mykje hjelp ein kan finne, vil likevel avhenge av kva delar av ordtilfanget det er snakk om, og kva type ordbok ein arbeider med. *Bokmålsordboka* og *Nynorskordboka* på den eine sida og *Norsk Ordbok* på den andre representerer to ulike ordboktypar både med omsyn til omfang, innhald og kor godt dokumentert eit lemma må vere for å takast med.

Standardordbøkene skal innehalde det mest sentrale og representative ordtilfanget i bokmål og nynorsk. Dei er dessutan primært skriftspråksordbøker. Artiklar med grunnlag i talespråket som allereie finst i ordbøkene, blir vanlegvis med vidare, men nye artiklar byggjer først og fremst på skriftlege belegg. Som vi har sett over, er talemålskorpusa små samanlikna med skriftspråskorpusa, og det er berre ein liten del av dei eksisterande oppslagsorda som er med i LIA-korpuset. Ei anna utfordring er at revisjonen skal vise fram endringar i språket sidan arbeidet med ordbøkene tok til for over 40 år sidan. Sidan størstedelen av LIA-materialet er eldre opptak, er korpuset mindre eigna. Det kan likevel vere til hjelp i revisjonen av enkeltartiklar, fordi det kan gje oss ein peikepinn om ord som er særleg vanlege i talemål, og ord som blir brukte litt annleis i tale enn i skrift.

Vi meiner LIA-korpuset er mest relevant for det vidare arbeidet med *Norsk Ordbok*. Verket skal dokumentere norsk både munnleg og skriftleg. Eit talemålskorpus er såleis ei nyttig kjelde. LIA-materialet har god geografisk spreining, og det er ein styrke med omsyn til ordbokarbeidet. *Norsk Ordbok* er ei dokumentasjonsordbok og har ikkje like strenge krav som standardordbøkene når det gjeld frekvens

og utbreiing. Også svakare dokumenterte ord frå målføra kan få ein plass der, og LIA-korpuset er difor ein ressurs for revisjonen av *Norsk Ordbok*.

## Referansar

- Bokmålsordboka*. 1986–. Språkrådet og Universitetet i Bergen. <https://ordbok.uib.no/>.
- Fjeld, Ruth Vatvedt. 2008. Talespråksforskningens betydning for leksikografien. I *Språk i Oslo: ny forskning omkring talespråk*, redigert av Janne Bondi Johannessen og Kristin Hagen, 15–28. Oslo: Novus.
- Hovdenak, Marit. 2013. Bokmålsordboka og Nynorskordboka gjennom ein generasjon. *Nordiske studier i leksikografi* 12, 229–246. Oslo: Nordisk forening for leksikografi.
- Innstilling om språksaken fra komitéen til å vurdere språksituasjonen m.v. oppnevnt ved kongelig resolusjon 31. januar 1964*. 1966. Oslo: Kirke- og undervisningsdepartementet.
- Kristiansen, Nina og Christian Emil Ore. 1998. *Sluttrapport Dokumentasjonsprosjektet*. Oslo: Dokumentasjonsprosjektet.
- Lyse, Gunn Inger. 2020. Ut med adamsslekt og inn med arveprinsesse? Leksikografiske metodar i revisjonen av Bokmålsordboka og Nynorskordboka. *Nordiske studier i leksikografi* 15, 215–224.
- 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*, redigert av Gisle Andersen, 31–50. Philadelphia: John Benjamins.
- Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet*. 1966–2016. Oslo: Det Norske Samlaget. <http://no2014.uib.no/>.
- Nynorskordboka*. 1986–. Språkrådet og Universitetet i Bergen. <https://ordbok.uib.no/>.
- Ordbokstatistikk*. 2020. Universitetet i Bergen. <https://ordbok.uib.no/perl/ordbokstatistikk.cgi>.

- Rauset, Margunn. 2019. Bokmålsordboka og Nynorskordboka – einnegga, toegga eller siamesiske tvillingar? *LexicoNordica* 26, 155–175.
- Språkrådet. 2015. *Retningslinjer for normering av bokmål og nynorsk*. <https://www.sprakradet.no/Spraka-vare/Norsk/normering/retningslinjer-for-normering/>.
- Søfteland, Åshild og Kaja Borthen. 2018. 'Æ e trønder, æ, sjø!' Den pragmatiske partikkelen 'sjø' i midt-norske dialektar. *Norsk Lingvistisk Tidsskrift* 36 (2), 249–280.
- Vikør, Lars S. og Åse Wetås. 2016. Norsk Ordbok: om folket – av folket – for folket. Historia til ordboksverket fram til i dag. I *Livet er æve, og evig er ordet. Festskrift til Norsk Ordbok 1930-2016*, redigert av Lars S. Vikør, Helene Urmland Karlsen og Åse Wetås, 15–39. Oslo: Det Norske Samlaget.
- Vikør, Lars Sigurdson. 2018. *Inn i Norsk Ordbok – Brukarrettleiing og dokumentasjon*. Oslo: Det Norske Samlaget. <http://no2014.uib.no/eNo/tekst/Inn%20i%20Norsk%20Ordbok.pdf>.
- Worren, Dagfinn. 1998. Om å avgrense eit ordtilfang – saga om målføreorda i Norsk handordbok. I *Normer og regler. Festskrift til Dag Gundersen 15. januar 1998*, redigert av Ruth Vatvedt Fjeld og Boye Wangenstein, 59–70. Oslo: Nordisk forening for leksikografi.

## English summary

In this article, we take a closer look at how a spoken language corpus like LIA Norwegian can be a resource in the task of revising the three Norwegian dictionaries *Bokmålsordboka*, *Nynorskordboka*, and *Norsk Ordbok*. One of the aims of the revision is finding new lemmas to include in the dictionaries and assessing and revising the dictionary entries which the dictionaries consist of at present. The tasks include deciding whether a word is a central part of modern vocabulary and therefore should be included, whether the definitions are in line with current language use, which usage examples are the most typical, and which are the most important collocations. It is necessary to have

good language sources, and in this article, we consider whether it is helpful or not to use the LIA corpus in the revision of the dictionaries.

Bente Selback  
Institutt for lingvistiske, litterære og estetiske studium  
Universitetet i Bergen  
bente.selback@uib.no

Terje Svardal  
Universitetsbiblioteket  
Universitetet i Bergen  
terje.svardal@uib.no