

Ut med *adamsslekt* og inn med *arveprinsesse*? Leksikografiske metodar i revisjonen av *Bokmålsordboka* og *Nynorskordboka*

Gunn Inger Lyse

In the project revising the two Norwegian dictionaries *Bokmålsordboka* and *Nynorskordboka* (www.uib.no/revisjonsprosjektet), the lexicographers are to revise these two existing dictionaries in tandem. This article presents the source material, tools, and methods that are utilised to enable the lexicographers to identify a modernised and relevant vocabulary in both dictionaries in an efficient and empirical manner.

1. Innleiing

Brukarvenlege søkjeverktøy og eit godt kjeldegrunnlag er avgjerande for at leksikografen kan arbeide effektivt og med solid empirisk forankring. Denne artikkelen presenterer kjeldematerialet, verktøy og metodar som blir brukte og utvikla i prosjektet Revisjon av *Bokmålsordboka* og *Nynorskordboka* (Revisjonsprosjektet). Dette prosjektet skal modernisere og forbetre dei digitale versjonane av ordbøkene.

Sentralt står dei vitenskaplege språksamlingane ved Universitetet i Bergen (UiB), som delvis ligg opne. I tillegg har prosjektet teke i bruk verktøy for språkkanalyse som er tilgjengelege via språkinfrastrukturen CLARINO (2019). Det eine verktøyet er Korpuskel, som let oss gjere avanserte søk på ord og frasar. Gjennom Revisjonsprosjektet er Korpuskel utvida slik at ein kan søkje i fleire korpus samstundes. Utvalet av korpus gjev til saman eit materiale på om lag 2,4 milliardar ord. Det andre verktøyet er INESS (2019), som let oss søkje etter syntaktiske konstruksjonar som ikkje er så lett å søkje etter i eit tradisjonelt tekstkorpus. Denne artikkelen gjev døme på når tekstkorpus er til god hjelp – men òg døme på at korpus kan vere misvisande.

Del 2 av artikkelen gjev ein kort introduksjon til Revisjonsprosjektet. I del 3 følgjer eit overblikk over kjeldegrunnlaget. I del 4 blir søkjesystema Korpuskel og INESS presenterte.

2. Revisjon av *Bokmålsordboka* og *Nynorskordboka*

Revisjonen av *Bokmålsordboka* og *Nynorskordboka* starta opp ved UiB i 2018 og er finansiert av Kulturdepartementet (sjå www.uib.no/revisjonsprosjektet). Ordbøkene inngår i Språksamlingane ved UiB, er eigd og drifta av UiB og Språkrådet, og dei ligg ope tilgjengeleg gjennom nettsøk og som nedlastbar app.¹

I Noreg står *Bokmålsordboka* og *Nynorskordboka* i ein særskild posisjon som Språkrådet sin viktigaste kanal for å formidle informasjon om gjeldande offisiell rettskriving til publikum (Språkrådet 2014:9). Desse er dei einaste fritt tilgjengelege digitale ordbøkene som gjer greie for heile breidda av normerte ord- og bøyingsformer, og dei blir løpande oppdaterte i tråd med rettskrivingsvedtak i Språkrådet. Ordbøkene er mykje brukte, med rundt 125 000 søk per døgn i 2018 (Rauset 2019:162).

Revisjonsprosjektet skal revidere dei to eksisterande ordbøkene parallelt, med ei tidsramme på fem og eit halvt år. Leksikografane har tre hovudoppgåver: sikre at definisjonane er i tråd med dagens språkbruk, peike ut eit modernisert og relevant ordtilfang i begge målformene, og gjere utvalet av lemma mellom ordbøkene likare (Selback 2020). Dette er omfattande produksjonsmål på avgrensa tid, og det er avgjerande med gode søkjeverktøy som let leksikografane jobbe etterretteleg og samstundes effektivt.

3. Kjeldegrunnlag

Bokmålsordboka og *Nynorskordboka* skal gje kvalitetssikra og oppdatert informasjon om normert rettskriving og om vanleg språkbruk for skriftspråka nynorsk og bokmål (Språkrådet & Universitetet i Bergen 2016). Som mellomstore ordbøker skal dei ha eit moderne ordtilfang. Den empiriske forankringa er viktig i ordbokarbeid både for å sikre kvaliteten på, og å etterprøve, arbeidet.

Eit tekstkorpus kan definerast som ei søkbar samling av digital tekst, der treffa kan bli analyserte på ulike måtar. Tekstkorpus er ei god empirisk kjelde både for å vurdere kva oppslagsord som bør vere med i ordbøkene (lemmautval), kva orda betyr, korleis orda oppfører seg syntaktisk og for å finne gode døme på typisk bruk av eit ord. Det er då vanleg å basere seg på ordfrekvens. Det er naturleg å ta omsyn til frekvensdistribusjonen over tid, altså om ordet er stabilt i bruk over tid, og korleis ordet er distribuert på tvers av domene, altså om ordet er del av allmennspråket. Sidan ordbøkene skal reflektere moderne skriftspråk, ser redaksjonen i hovudsak på ordfrekvensen innanfor dei siste 50 åra.

¹ Ordbøkene på nett: ordbok.uib.no; appen heiter Ordbøkene.

Ordbokarbeidet kviler likevel ikkje åleine på søk i tekstkorpus. For det første tek vi omsyn til normeringsvedtak i Språkrådet. Ordbokredaksjonen ved UiB vel ut kva slags ord og tydingar som skal vere med i ordbøkene, medan Språkrådet bestemmer korleis orda skal skrivast og bøyast. For det andre bruker vi dei vitenskaplege Språksamlingane ved UiB, som ligg ope tilgjengeleg (www.usd.uib.no). Mellom anna sjekkar vi i Setelarkivet og dei andre kjeldesamlingane til *Norsk Ordbok* det som vi ikkje finn belegg for i tekstkorpus. For det tredje konsulterer ordbokredaktøren òg språkkjensla si og eigen grammatikkunnskap i vurderinga av korleis ei leksikografisk eining blir brukt.

4. Søkjeverktøy for språkanalyse

Generelt er det vanskeleg å finne korpus som let deg gjere søk både ut frå frekvens, tid og domene, samstundes som materialet er stort nok (Gjesdal & Lyse 2016). Forskingsinfrastrukturar som CLARIN er interessante, fordi dei jobbar for å gjere ulike korpus tilgjengelege og søkbare gjennom felles grensesnitt (CLARIN 2019). Gode døme på dette er Glossa (Nøklestad et al. 2017), Sketch Engine (Kilgarriff et al. 2014), Korp (Borin et al. 2012), Korpuskel (Meurer 2012) og Weblicht (Hinrichs et al. 2010). Nasjonalbiblioteket i Noreg utviklar eit programmeringsgrensesnitt (ein API) for å gjere språkanalysar på alt det digitaliserte materialet som ligg hos Nasjonalbiblioteket (Johnsen 2019).

Ordbokredaktørane i Revisjonsprosjektet brukar i dag hovudsakleg to verktøy for språkanalyse som er tilgjengelige gjennom den språkteknologiske infrastrukturen CLARINO (2019). Desse to verktøya blir presenterte nedanfor.

4.1. Korpuskel-Lex

Korpuskel er eit grensesnitt for å tilgjengeleggjere, søkje i, og analysere, tekstkorpus (Meurer 2012). Korpuskel er eit generisk verktøy med over 60 korpus på ulike språk, format og mediatypar. Korpuskel-Lex er ei vidareutvikling av Korpuskel som vart implementert i 2018 for Revisjonsprosjektet, og let brukaren søkje på tvers av mange korpus samstundes. Som del av CLARINO er Korpuskel-Lex ope tilgjengeleg, innanfor rammene av brukslisensane som gjeld for dei enkelte tekstkorpusa.

Korpusa i Korpuskel-Lex utgjer om lag 2,4 milliardar ord, og eit oversyn over korpusa ligg føre i tabell 1. Dette utvalet svarar på fleire behov. For det første er dette korpus der vi kan vente at teksten har vore gjennom ein redaksjonell prosess før publisering (i motsetnad til t.d. bloggtekst). Dette er viktig for

å ha eit kjeldegrunnlag av ein viss skriftleg kvalitet. Revisjonsprosjektet treng vidare å skilje mellom bokmål ('bm') og nynorsk ('nn') (jf. kolonna *målform*). Det er nyttig med eit stort materiale frå ulike domene (jf. kolonnene *storleik* og *sjanger*) og som dekkjer eit breitt tidsaspekt (jf. kolonna *tidsrom*) – dette fordi vi rett nok fokuserer på moderne språk, men samstundes treng vi i revisjonen òg å sjå bruken av orda i eit større tidsperspektiv. Vi ønskjer å kunne søkje på språkleg analysert materiale (lemma og ordklasse) der det er mogleg (jf. kolonna *søkjeattributt*).

Korpusnamn	Målform	Storleik (mill.)	Tidsrom	Sjanger	Søkjeattributt
Aviskorpus (bokmål)	bm	1880,62	1998–2019	sakprosa (avis)	ordform
Aviskorpus (nynorsk)	nn	19,6	1998–2015	sakprosa (avis)	ordform
Aviskorpus annotert	bm	28,97	2001–2009	sakprosa (avis)	ordform, lemma, ordklasse
Forskning.no (2017)	bm	21,47	1998–2017	sakprosa (avis)	ordform, lemma, ordklasse
Leksikografisk bokmålskorpus	bm	102,29	1985–2013	blanda (sakprosa, skjønnlitteratur)	ordform, lemma, ordklasse
NBs frie tekster (bokmål)	bm	516,39	1765–2013	blanda (sakprosa, skjønnlitteratur)	ordform
NBs frie tekstar (nynorsk)	nn	46,02	1850–2010	blanda (sakprosa, skjønnlitteratur)	ordform
Norsk Ordboks nynorskkorpus	nn	107,8	1866–2012	blanda (sakprosa, skjønnlitteratur)	ordform, lemma, ordklasse
Talk Of Norway	bm, nn	63,8	1999–2016	sakprosa	ordform, lemma, ordklasse

Tabell 1: Tekstkorpusa som er del av Korpuskel-Lex.

Korpuskel-Lex har alle funksjonar som òg finst i Korpuskel, i tillegg til to funksjonar som har blitt utvikla særskilt for behova til ordbokredaktørane. Den første er at Korpuskel-Lex, til skilnad frå Korpuskel, let brukaren søkje i meir enn eitt korpus om gongen. Treffa kan bli sorterte ut frå ulike attributt som korpuskjelde eller tidsrommet treffa er fordelt over (jf. kolonnane i figur 1). Dermed får redaktøren raskt overblikk over bruken av eit ord, utan å måtte gå frå kjelde til kjelde og sjekke.

corpus		treff		source	year
avis-nno	s og Medkila i Prestegardsmoen i dag. Ballen	curla	inn i lengste hjørnet utan sjanse for målvakt	BT	2008
avis-nno	det meste sjølv då han skar inn frå høgre og	curla	ballen i motsett hjørne. 0-2, og alt låg til ret	AP	2011
avis-nno	det meste sjølv då han skar inn frå høgre og	curla	ballen i motsett hjørne. 0-2, og alt låg til ret	SA	2011
avis-nno	det meste sjølv då han skar inn frå høgre og	curla	ballen i motsett hjørne. 0-2, og alt låg til ret	AA	2011
avis-nno	det meste sjølv då han skar inn frå høgre og	curla	ballen i motsett hjørne. 0-2, og alt låg til ret	FV	2011
avis-nno	r kjøpt OL-rettar. Altså risikerer vi i framtida	curling	med reklameavbrot. Samtidig har NRK vist a	AP	2011
avis-nno	all, tennis, basketball, bordtennis, volleyball,	curling	, golf og synkronsymjing. Kvifor akkurat des	BT	2015
avis-nno	YTO: ARNE S. GJONE ¶ Les også En skole for	curlingbarna	- 06.02.2005 Svikt i lærebøkene, ¶ svikt hos	AP	2005
avis-nno	ar det, går det ut på eitt. Nokre har kalla dei	curlingbarna	, barna som får kosta banen sin fri av foreldr	BT	2006
avis-nno	nge menneske i dag er ikkje kalla glasur- og	curlinggenerasjon	for ingenting, men det er det vaksne som ha	SA	2015
avis-nno	lange kallenamn, som «lydig», «flinkis» og «	curlinggenerasjonen	». Angrepa kjem gjerne når nokon går ut i rr	BT	2014
avis-nno	lt; F ¶ ~ Norsk seier over Japan ¶ De norske	curlinggutta	slo Japan 8-5 i torsdagens første VM-kamp. '	SA	2006
avis-nno	det på havets botn ligg ein tusen år gammal	curlingstein	. -  Nei, nei - han er ikkje tusen - ha	BT	2015
avis-plain	Paramount går snart i gang med " Girl in the	curl	", og Warner Bros skal lage " Lisa Andersen :	SA	2001

Figur 1: Skjermbilete frå Korpuskel-Lex: konkordansar for søk etter ord som byrjar på *curl*.

Den andre funksjonen er at det er mogleg å avgrense søkjetreffa ut frå ord i konteksten. I *Nynorskordboka* er t.d. ordet *appelsinbåt* oppført med ei generell tyding ('bit av ein skrelt appelsin') og den meir spesifikke tydinga 'båt som fraktar appelsin', som stammar frå tida då appelsinar vart frakta til Noreg med båt. For å finne korpusdøme på denne meir spesifikke tydinga, kan ein raffinere søket til at ordet *kome* må finnast innanfor ei ramme på t.d. 10 ord før og etter *appelsinbåt* (som i: *appelsinbåten kom*).

Korpuskel-Lex har eit avansert søkjespråk med regulære uttrykk, og har òg eit forenkla, grafisk grensesnitt for søk (Meurer 2019). Det er mogleg å lagre ofte brukte søk, enten privat for den innlogga brukaren eller ope, slik at alle i ordbokredaksjonen kan dele søk. Ein kan òg spare søket som ein URL, som let deg repetere same søk i dei same korpusa i etterkant.

Der er fleksible moglegheiter for å vise resultatet etter eit søk. Treffa kan bli viste som konkordansar (vising av søkjeordet i kontekst), som i figur 1, eller som frekvensordlister, som i figur 2. Frekvensordlister, som i figur 2, er praktiske for å peike ut gode kandidatar til oppslagsformer som bør vurderast med i ordboka. Døme på dette er ordet *curlingforeldre* (nedst i figur 2), som ikkje låg i ordboka allereie. Lister som dette viser òg i kva grad eit ord er produktivt i samansetjingar, som i enkelte høve kan gjere at eit grunnord vert teke med, sjølv om ordet òg er spesielt frekvent.

Antall treff: 6693, unique values: 429. Attributt: word | ignorer storskriving | sorter: etter frekve

Side 1 av 2. Previous Next

1514 (22,62%)	curling	21 (0,31%)	curlingpresident	7 (0,10%)	curling-
913 (13,64%)	curling-VM	20 (0,30%)	curlinglandslag	7 (0,10%)	curling-gull
439 (6,56%)	curlinggutta	20 (0,30%)	curlingspilleren	7 (0,10%)	curlingbarn
301 (4,50%)	curlet	19 (0,28%)	curlingguttene	7 (0,10%)	curlingbarna
204 (3,05%)	curling-EM	18 (0,27%)	curlingdamene	7 (0,10%)	curlingdamer
159 (2,38%)	curlingjentene	18 (0,27%)	curlingsteiner	7 (0,10%)	curlingforelder
150 (2,24%)	curlinglaget	16 (0,24%)	curlingforbundets	7 (0,10%)	curlingkarrieren
129 (1,93%)	curlinglandslaget	16 (0,24%)	curlingklovnene	7 (0,10%)	curlingkolleger
121 (1,81%)	curlinghallen	15 (0,22%)	curling-gutta	7 (0,10%)	curlinglagene
116 (1,73%)	curlingherrene	15 (0,22%)	curlingklubb	7 (0,10%)	curlingnasjon
113 (1,69%)	curlinglag	14 (0,21%)	curlingbukse	6 (0,09%)	curling-finalen
112 (1,67%)	curle	14 (0,21%)	curlinggull	6 (0,09%)	curlingdrønning
111 (1,66%)	curler	14 (0,21%)	curlingkvinner	6 (0,09%)	curlingfans
99 (1,48%)	curlingmillioner	14 (0,21%)	curlingmiljøet	6 (0,09%)	curlingfeberen
71 (1,06%)	curlingforbundet	14 (0,21%)	curlingseier	6 (0,09%)	curlingfolket
65 (0,97%)	curlingbanen	13 (0,19%)	curlere	6 (0,09%)	curlinghistorie
62 (0,93%)	curlingguttas	12 (0,18%)	curla	6 (0,09%)	curlinginteressen
60 (0,90%)	curlingforeldre	12 (0,18%)	curlingbukser	6 (0,09%)	curlingjenter
54 (0,81%)	curling-seiser	12 (0,18%)	curlingfeber	6 (0,09%)	curlingkampene

Figur 2: Skjermbilete frå Korpuskel-Lex: frekvensordliste, som resultat på søk etter ord som byrjar på *curl*.

Treffa kan òg bli visualiserte med distribusjonen over tid og på tvers av korpus (trendlinjer), som er nyttig for å vurdere kor etablerte orda er i språket. Som døme kan vi ta orda *adamsslekt* ('menneska'), som ligg i begge ordbøkene i dag, og *arveprinsesse*, som ikkje låg i ordbøkene då Revisjonsprosjektet tok til. I våre kjelder fann vi berre tre treff på *adamsslekt* som er nyare enn 1940 (eitt frå 1971 og to frå 1981). Til samanlikning har *arveprinsesse* ein ganske stabil frekvens ut over 2000-talet. Korpuskel-Lex viser òg enkle statistiske kollokasjonar, slik at eit søk på *hatt* kan returnere statistisk signifikante kollokasjonar som *fjør i hatten*.

Sjølv om korpus og frekvensar er nyttige leksikografiske reiskaper, må ein hugse på at frekvensar ikkje fortel alt. Eit ord kan tilsynelatande ha høg frekvens, samstundes som frekvensen kan vere ujamnt distribuert over tid eller på tvers av ulike korpus eller domene. Korpuskel-Lex har nok dessutan, per i dag, ei uønskt overvekt av avistekst.

Eit døme på ujamn distribusjon er ordet *aortastenose* ('hjerterklaffproblem'). Dette ordet kom opp medan redaktørane jobbe med bokstaven *a* og søkte på ord som byrja på *aor-*. Ordet *aortastenose* stod for 51,33 prosent av alle treffa i resultatlista (totalt 113 treff), medan det intuitivt meir kjende ordet *aorta* berre var halvparten så frekvent. Men ser ein på distribusjonen til treffa på *aortastenose* (jf. figur 3), ser ein at nesten alle treffa er frå 2005 – som var då norske kong Harald vart innlagd på sjukehus med hjerterklaffproblem. Dette ordet fekk såleis ein relativt høg frekvens i eit veldig kort tidsrom, og hovudsakleg i avistekst. Liknande døme er ord som vart svært frekvente i ein kort periode i samband med vulkanutbrotet på Island i 2010 (t.d. *askesky*).

	(sum)	<input checked="" type="checkbox"/> aortastenose
(sum)	58 (100,0)	58 (100,0)
	58 (100,0)	58 (100,0)
2005	57 (100,0)	57 (100,0)
2014	1 (100,0)	1 (100,0)

Figur 3: Skjermbilete frå Korpuskel-Lex: distribusjonen av ordet *aortastenose* relativt til år.

På same måte kan eit ord vise seg å dominere i eit bestemt korpus eller domene – t.d. har ordet *assistenttrener* over 13 000 treff i dei samla korpusa som er tilgjengeleg i Korpuskel-Lex. Men ser ein på distribusjonen per korpus, ser ein at dette ordet er overrepresentert i Aviskorpuset (12 977 av treffa), som har veldig mykje tekst frå domenet sport. Ordet vart likevel inkludert som eit nytt ord i ordbøkene, ettersom ein kan argumentere for at høgfrekvente idrettsomgrep kan reknast som ein del av allmennspråket. Redaksjonen må difor aktivt undersøkje kvar treffa er funne, og òg konsultere andre kjelder enn dei primære tekstkorpusa som vi bruker.

4.2. NorGramBank i INESS

Det andre verktøyet vi har teke i bruk er den norske trebanken NorGramBank, som er søkbar gjennom verktøyet INESS (Rosén et al. 2012). Ein trebank kan definerast som eit tekstkorpus der kvar setning har ein detaljert syntaktisk analyse. INESS er ein infrastruktur for å bevare og gjere trebankar tilgjengelege for brukarar, og er del av CLARINO. NorGramBank er ein trebank for bokmål og nynorsk (Dyvik et al. 2016). Analysene i denne trebanken er produserte automatisk med bruk av den maskinelle grammatikken NorGram, som er skriva innanfor rammeverket leksikalsk-funksjonell grammatikk (LFG). NorGramBank inneheld om lag 80 millionar ord, fordelt på skjønnlitteratur, sakprosa og stortingstekst. Fordelinga mellom bokmål og nynorsk i NorGramBank er som i materialet i Korpuskel-Lex, med om lag 90 prosent bokmål og 10 prosent nynorsk.

NorGramBank let ordbokredaktøren søkje opp WordSketch-liknande informasjon om ord, som dei vanlegaste verbrammene til eit ord (*gifte seg, gifte bort, gifte seg til*) eller typiske døme på korleis at-setningar blir styrte av preposisjonar (*etter at, utan at*). NorGramBank har vist seg særleg nyttig for funksjonsord (fordi ein kan gjere søk baserte på dei ulike funksjonane til ordet), ord med systematiske skilnader knytt til syntaktiske relasjonar (t.d. partisippformer og adjektiv) og fleirordsuttrykk (t.d. *auge: ha auge for, få auge på, falle i auga*). Fi-

gur 4 viser døme på søk etter fleirordsuttrykk som inneheld bokmålssubstantivet *øye*. Eit fleirordsuttrykk treng berre ein forekomst i INESS for å bli registrert (sjå Dyvik et al. 2019). Leksikografen må altså framleis vurdere om fleirordsuttrykka er tilstrekkeleg etablerte i allmennspråket til å bli tekne med i ordboka.

Som dømet i figur 4 viser, vart det funne tre ulike fleirordsuttrykk i NorGramBank som inneheld *øye* (*få øye på*, *holde øye med*, *ha øye til*), sorterte etter frekvens (kolonna lengst til venstre). Ved å klikke på eit fleirordsuttrykk, t.d. på *holde#øye*med*, kan ein gjennomgå setningane som fleirordsuttrykket vart funne i.

Click on a row to see the matching sentences. |
Copy format: plain BRO

Count	#p: <i>atom</i>	globals: <i>language</i>
76	få#øye*på	nob
30	holde#øye*med	nob
1	ha#øye*til	nob

Figur 4: Skjermbilete frå INESS: fleirordsuttrykk som er registrerte i den norske trebanken NorGramBank for substantivet *øye* på bokmål, sorterte etter frekvens.

5. Konklusjon

Korpus er til god hjelp for leksikografen når det gjeld lemmatilfang, fleirordsuttrykk, ordtydingar, syntaktisk åtferd og bruksdøme. Samstundes må korpus brukast klokt, og dei kan ikkje erstatte leksikografiske vurderingar. I Revisjonsprosjektet ser vi ein stor gevinst i å knytte IT- og datalingvistisk kompetanse til leksikografiske prosjekt som driv korpusutvikling og -tilpassing etter behova som ulike prosjekt har. Vi finn stor synergieffekt i å nytte aktivt andre forskingsprosjekt og infrastrukturar ved UiB – fordi aktiv bruk frå vår side stimulerer til betre løysingar som òg kjem nye brukarar til gode. T.d. har korpusverktøyet Korpuskel blitt utvida til å få nye tekniske løysingar som er tilgjengeleg for fleire, grunna våre behov. Samstundes ser vi behovet for å vidareutvikle både tekniske løysingar og statistiske hjelpemiddel som kan forenkle og effektivisere ordbokarbeidet ytterlegare.

Litteratur

Ordbøker

Bokmålsordboka og Nynorskordboka = Universitetet i Bergen og Språkrådet.
<ordbok.uib.no> (oktober 2019)

Annan litteratur

Borin, Lars, Markus Forsberg & Johan Roxendal (2012): Korp – the corpus infrastructure of Språkbanken. I: *Proceedings of LREC 2012*. Istanbul: ELRA, 474–478.

CLARIN (2019): Nettside for den europeiske infrastrukturen CLARIN. <www.clarin.eu> (oktober 2019).

CLARINO (2019): Nettside for CLARINO, den norske noden i CLARIN. <clarin.w.uib.no> (oktober 2019).

Dyvik, Helge, Gyri Smørdal Losnegaard & Victoria Rosén (2019): Multiword expressions in an LFG grammar for Norwegian. I: *Representation and parsing of multiword expressions: Current trends*. Language Science Press, 69–108.

Dyvik, Helge, Paul Meurer, Victoria Rosén, Koenraad De Smedt, Petter Haugereid, Gyri Smørdal Losnegaard, Gunn Inger Lyse & Martha Thunes (2016): NorGramBank: A ‘Deep’ Treebank for Norwegian. I: Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asunción Moreno, Jan Odijk & Stelios Piperidis (red.): *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*. Portorož, Slovenia: ELRA, 3555–3562.

Gjesdal, Anje Müller & Gunn Inger Lyse (2016): Exploring an environmental neologism in Norwegian across corpora. I: *Neologica* 10, 39–57.

Hinrichs, Erhard W, Marie Hinrichs & Thomas Zastrow (2010): WebLicht: Web-Based LRT Services for German. I: *Proceedings of the ACL 2010 System Demonstrations*, 25–29.

INESS (2019): Nettside for infrastrukturen INESS. <clarino.uib.no/iness> (oktober 2019).

Johnsen, Lars G. (2019): *Mining digital libraries*. 23. april 2019. <www.github.com/Yoonsen/Tutorial-DHN2019> (oktober 2019).

Kilgarrieff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel (2014): The Sketch Engine: ten years on. I: *Lexicography*. Springer, 7–36.

- Meurer, Paul (2012): Corpuscle – a new corpus management platform for annotated corpora. I: Gisle Andersen (red.): *Exploring Newspaper Language: Using the Web to Create and Investigate a large corpus of modern Norwegian*. Studies in Corpus Linguistics 49, John Benjamins.
- Meurer, Paul (2019): *Corpuscle-Lex: The query language*. <clarino.uib.no/lex/page?page-id=The_query_language> (oktober 2019).
- Nøklestad, Anders, Kristin Hagen, Janne Bondi Johannessen, Michal Kosek & Joel Priestley (2017): A modernised version of the Glossa corpus search system. I: Jörg Tiedemann (red.): *Proceedings of the 21st Nordic Conference on Computational Linguistics (NoDaLiDa)*, 251–254.
- Rauset, Margunn (2019): Bokmålsordboka og Nynorskordboka – einegga, toegga eller siamesiske tvillinger? I: *LexicoNordica* 26, 155–175.
- Rosén, Victoria, Koenraad De Smedt, Paul Meurer & Helge Dyvik (2012): 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*. Istanbul, Turkey: ELRA, 22–29.
- Selback, Bente (2020): «Å nei, det ordet er ikkje lov på nynorsk! » Eller ...? Om parallell redigering av to norske ordbøker. I: *Nordiska studier i lexikografi* 15, 297–305.
- Språkrådet (2014): *Norsk ordbokpolitikk*. Notat fra Språkrådet til Kulturdepartementet 3. mars 2014. <www.sprakradet.no/upload/Aktuelt/Notat%20om%20norsk%20ordbokpolitikk.pdf> (mars 2019).
- Språkrådet & Universitetet i Bergen (2016): *Samarbeidsavtale*. Upublisert.

Gunn Inger Lyse
 ordbokredaktør, ph.d.
 Universitetet i Bergen
 Sydneplass 7
 NO-5007 Bergen
 gunn.lyse@uib.no