

NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Et fullformsystem for analyse av eldre tekst på tidlig nynorsk, bygd på Aasen-normalen	
Forfatter:	Christian-Emil Smith Ore, Oddrun Grønvik & Trond Minde	
Kilde:	Nordiska studier i lexicografi 16, 2023, s. 267–279	
URL:	https://tidsskrift.dk/nsil/issue/archive	

© Respektive författare, Nordiska föreningen för lexicografi och Meijerbergs institut för svensk etymologisk forskning, 2023

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavspersonen til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Et fullformsystem for analyse av eldre tekst på tidlig nynorsk, bygd på Aasen-normalen

Christian-Emil Smith Ore, Oddrun Grønvik & Trond Minde

Many European languages have undergone considerable changes in orthography over the last 150 years. This hampers the application of modern computer-based analysers to older text, and hence computer-based annotation and studies of text collections spanning a long period. In 2021, as a step towards a functional analyser for Norwegian texts (Nynorsk standard) from the 19th century, we did a pilot project with the objectives to create a full form generator for all inflected forms of headwords found in Ivar Aasen's dictionary published in 1873 (Aasen 1873) and based on the description in his grammar from 1864 (Aasen 1864).

As a test, the full form list generated from this new word bank was used to analyse the word inventory of texts by A.O. Vinje, written in the period 1850–1870. The Vinje texts were also analysed using a full form list of modern standard Norwegian, to study the differences in applicability and see how Vinje's language relates to the written standard of modern Norwegian.

KEYWORDS: digital lexicography, full form systems, 19th century, orthographic history, text analysis

1. Bakgrunn

Norsk Ordbok har som andre store nasjonalordbøker vært et langvarig prosjekt. Det ble påbegynt rundt 1935 og redigeringen ble avsluttet i 2015. I 2002 ble det norske kulturdepartementet og Universitetet i Oslo enige om en restrukturering av prosjektet med nødvendig finansiering under forutsetning av at ordboka ble ferdigstilt til grunnlovsjubileet i 2014. Det holdt nesten, og ordboka forelå i 12 bind i mars 2016. I dette prosjektet, også kalt NO2014, redigerte en alfabetdel I–Å og halve H. Det er i de senere årene satt i gang et prosjekt for å revidere A–H. En av forutsetningene for NO2014-prosjektet var at ordboka skulle redigeres digitalt på en slik måte at den også kunne publiseres som en digital ordbok sammen med det viktigste grunnlagsmaterialet. Store deler av dette grunnlagsmaterialet, blant annet seddelsamlingen på 3,2 millioner sedler, var allerede

digitalisert på 1990-tallet (Ore 1998). Det digitale materialet ble koplet til redigeringsdatabasen via Metaordboka (Ore 1998, 2012; Ore & Grønvik 2018), som er et verktøy for å systematisere leksikografisk kildemateriale og er spesielt egnet for dokumentasjon av svakt normerte språk som dialekter og språk med stor ortografisk variasjon. I tillegg til Norsk Ordboks seddelsamlinger og nyordsamlingen for bokmål inneholder Metaordboka et digitalt ordbokbibliotek, med 95 store og små ordbøker.

Metaordboka er et register over det norske leksikonet der hver artikkel representerer en leksikalsk gjenstand (Atkins & Rundell 2008), utvidet til å omfatte ortografiske varianter. I tillegg til bruksbeleggene har en metaordboksartikkel en oppslagsdel med redaksjonelle varianter. For hver variant er språk (bokmål/nynorsk), ordklasse, rettskrivningsstatus og tidsrom for denne statusen angitt. I tillegg er sammensetninger ledanalyisert. Det er for tiden om lag 780 000 artikler i Metaordboka. Vi har i de siste seks årene sett på hvordan den kan brukes til dokumentasjon og analyse av skriftspråkutviklingen i Norge de siste 150 årene. Et første mål i dette arbeidet var å se i hvilken grad vi kunne bruke kildene i Metaordboka til å samordne ordtilfanget i skriftmålsstandardene bokmål og nynorsk (Ore & Grønvik 2018a), og deretter å se om ordbokbiblioteket i Metaordboka viste utviklingen av et kjernevokabular for bokmål (Ore & Grønvik 2018b). Siden en rekke offisielt godkjente ordlister og ordbøker er knyttet til Metaordboka, danner den et skjelett for en historisk normordbok. *Det Centrale Ordregister*, som ble presentert på NFL 2022 (Henrichsen 2023), har mye av den samme funksjonaliteten som Metaordboka, men har et litt annet fokus, struktur og implementasjon. Men det kan være interessant å se på de to i sammenheng siden norsk bokmål og dansk har en felles historie til rundt 1900.

Alle ressursene og infrastrukturen som er nevnt i denne artikkelen, vedlikeholdes nå av *Språksamlingane* ved Universitetet i Bergen. En del av resultatene i denne artikkelen er også beskrevet i Ore, Grønvik & Minde (2022).

2. Norsk ordbank

I tillegg til Metaordboka kommer Norsk ordbank, som er en oversikt over grunnord og paradigmer samt fullformgeneratorer for bokmål og for nynorsk. Den bygger på IBMs stavekontroll fra 1980-tallet og arbeid

gjort i forbindelse med et prosjekt for å lage en morfo-syntaktisk, regelbasert tonivåtagger (Koskeniemi 1990) for moderne norsk (Hagen et al. 2000). Grunnordene i ordbanken er lenket til oppslagsordene i Bokmålsordboka (Wangensteen 1986–2006, Ordbøkene 2023) og Nynorskordboka (Hovdenak et al. 1986–2006, Ordbøkene 2023). Oppslagsord og bøyingsinformasjon i ordbøkene hentes direkte fra ordbanken (se <https://ordbokene.no>), og Språkrådets anbefalte rettskrivning formidles via ordbanken.

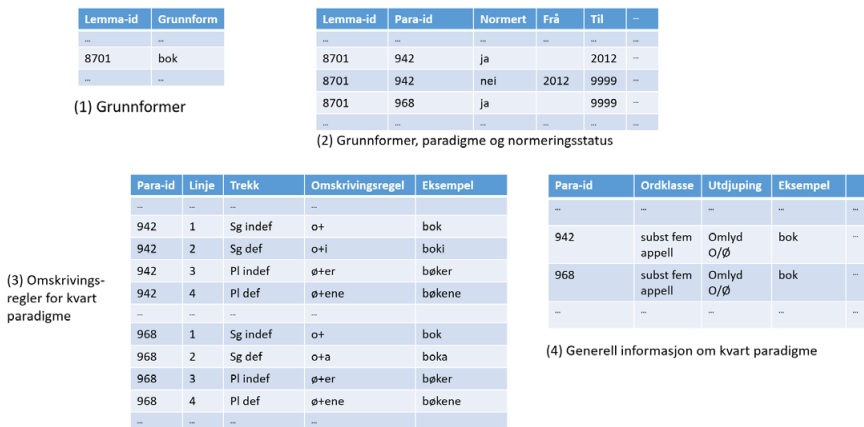
I vårt arbeid forstår vi med en ordbank en datastruktur for å lagre informasjon om ord (leksikalske gjenstander) og deres bøyde former. Den grunnleggende ideen er at et ord (leksikalsk gjenstand) er identifisert som mengden av alle dets mulige bøyde former. I denne modellen vil oppslagsformen i en ordbok kunne brukes som en representant for hele mengden av bøyde former. Både for bokmål og nynorsk har det vært mange ganske gjennomgripende revisjoner siden slutten av attenhundretallet.

Figur 1 viser en forenklet versjon av en ordbank. Strukturen består av (1) en liste av grunnformer (oppslagsord), (2) en tabell med informasjon om hvilket eller hvilke paradigmer grunnformen kan følge, og (3) en liste av omskrivningsregler (paradigmer) for å generere de bøyde formene fra grunnformen. Tabellen som vises i (2), er krumtappen i systemet. For en gitt grunnform vil en linje i tabellen fortelle hvilket eller hvilke paradigmer grunnformen følger, den ortografiske status og tidsrommet for denne. I nynorsk hadde for eksempel ordet 'bok' inntil 2012 sideformen 'boki' i tillegg til 'boka' i bestemt form entall. Siden 2012 har formen 'boka' vært eneste tillatte form. Paradigmatabelen (3) i figur 1 viser noen av de omskrivningsreglene som brukes for å lage fullformene. Omskrivningsprosessen fungerer som følgende: En grunnform sammenliknes med omskrivningsreglen i linje 1. I eksempelet i figur 1 vil jokertegnet '+' bli bundet til 'k'. Det vil si at det konstante initialsegmentet vil være 'b', mens finalsegmentet som endres, er 'ok'. Linjene 2 til 4 angir endringene som må gjøres for å lage fullformlistene 'bok, boki, bøker, bøkene' og 'bok, boka, bøker, bøkene'. Det kan være stort overlapp mellom fullformlistene generert av ulike paradigmer for ett og samme ord slik eksempelkolonnen viser. I en morfologisk analysator vil en som oftest være interessert i å finne grunnformen (lemma) og hvilken paradigmelinje (morfologiske trekk) formen representerer. En vil derfor slå sammen slike overlapp og lage en liste uten duplikater av parene (fullform, morfologiske trekk). Sys-

temet med omskrivningsregler er basert på en stavekontroll utviklet av IBM på slutten av 1980-tallet (Engh 2014).

Hvert paradigme har i tillegg til det som er vist i figur 1, informasjon om ordklasse, prototypiske grunnformer og eventuelle kommentarer. I Aasen-ordbanken har en for hvert paradigme referanser til de relevante paragrafene i Aasens grammatikk (Aasen 1864).

En tilleggsgevinst er at ordbankene dokumenterer endringene i rettskrivningen foreløpig siden 1995. Siden all kopling mellom grunnord og paradigmer er angitt med normeringsstatus og tidsrom, kan en hente ut rettskrivningsnormen slik den var på et gitt tidspunkt.



FIGUR 1 En skisse av strukturen i ordbanken for moderne nynorsk.

Ordbanken og Metaordboka har ulik historie og formål, men begge dokumenterer det norske ordinventaret. Ordbankens fokus er moderne norsk slik vi finner det i Bokmålsordboka og i Nynorskordboka. Metaordboka favner mye videre og dokumenterer ordinventaret tilbake til slutten av 1800-tallet. Ordbanken inneholder en fullformgenerator og genererer alle mulige bøyde former og markerer hvilke som er i samsvar med Språkrådets anbefalinger. En slik funksjon finnes ikke i Metaordboka.

3. Analyseverktøy for eldre tekst

Norsk ordbank er begrenset til moderne rettskrivning, og en analysator basert på ordbanken fungerer ikke særlig godt for tekster fra før de store

rettskrivningsreformene på 1930-tallet. Om en kunne utvide fullformgeneratorene til å omfatte nå foreldede bøyingsformer av grunnord i eldre rettskrivning, vil en kunne bygge analysatorer med bedre dekningsgrad. Det beste ville være å lage fullformer for alle oppslagsord i Metaordboka, men det reiser prinsipielle spørsmål. Om en skal lage en analysator for eldre tekst, må eldre ordformer kunne bøyes etter en antatt offisiell standard på tidspunktet for belegget, og hvordan skal en (re)konstruere eldre standarder, som vil ha hypotetiske former?

En vanlig strategi for å konstruere en hypotetisk standard er å gjennomføre en analyse av utvalgte tekster fra den perioden en vil arbeide med. På grunnlag av denne lager en så en oversikt over leksikalske gjenstander, hvordan de er realisert, og etablerer en tentativ norm med grunnformer og bøyingsformer. Dette er en standard metode for en leksikografisk beskrivelse av språk. En annen vei, som forutsetter at språket til en viss grad er dokumentert, er å bruke relevante, autoriserte ordlister og ordbøker som ligger nærmest opp til (helst før) den perioden en vil studere, og å bruke grammatikker fra samme periode til å sette opp bøyingsskjema. Denne metoden er godt egnet til å se i hvilken grad en (konstruert) norm er brukt i tekster fra den gitte perioden. Men et problem er at slike ordlister ofte har relativt få oppslagsord og dermed har en lav dekningsgrad. Dette er typisk for germanske språk med produktiv ordsammensetning. Den beste løsningen er å kombinere de to tilnærmingene, altså både analysere et korpus av eldre tekster og ta utgangspunkt i tilgjengelige grammatikker og ordlister. Begge krever mye arbeid, men kan gi et nyttig analyseredskap.

Som et første ledd i arbeidet med analysatorer for eldre bokmål og nynorsk har vi gjennomført et fullskala pilotprosjekt der vi tok utgangspunkt i det som danner begynnelsen for nynorsk (landsmål) som skriftspråkstandard, nemlig Ivar Aasens *Norsk Grammatik* (Aasen 1864) og *Norsk Ordbok med dansk Forklaring* (Aasen 1873). Til sammen danner de en beskrivelse av norsk folkemål og et forslag til en norsk rettskrivning.

Aasen var selvlært lingvist i en tid med mange selvlærte lingvister og en raskt voksende sammenlignende historisk språkvitenskap. Aasen studerte den danske lingvisten Rasmus Rasks arbeider og ulike språk som norrønt, latin, gresk og moderne språk. I 1840 fikk Aasen i oppdrag fra *Det Kongelige Norske Videnskabers Selskab* (DKNVS, Trondheim) å dokumentere det norske folkespråket gjennom dialektene og teste hypotesen om at norsk språk stammet fra norrønt og ikke var et forvansket dansk (Aasen

1848 og 1850, Walton 2016). Å dokumentere norsk folkespråk var en oppgave han fortsatte med resten av livet. En skal være klar over at han skrev for forskersamfunnet, og at han vurderte å skrive ordforklaringene i ordbøkene på tysk for å nå et større publikum, men endte med danske ordforklaringer. Formålet var altså å dokumentere norsk folkemål og ikke å lage en ordbok til støtte i daglig tekstproduksjon. Det hadde konsekvenser for ordutvalget i ordboka. Nyere importord ble stort sett utelatt. Vi har gitt en bredere beskrivelse av Ivar Aasen og hans ordbok og grammatikk (Aasen 1873 og Aasen 1864) i det andre bidraget på NFL 2022 (Grønvik, Ore & Minde 2023).

Grunnlagsmaterialet for pilotprosjektet er Aasens ordbok (1873) og grammatikk (Aasen 1864). Ordboka ble skrevet inn som formatert digital tekst i 1996. Den ble publisert i en CD-utgivelse og senere på Nynorsk kultursentrums nettsider, og var grunnlaget for Aarset og Krukens trykte versjon fra 2003 (Aasen 2003). I 2016 ble den digitale teksten (semi-)maskinelt analysert og gitt en TEI-koding, se Ore (2016). Vi ønsket bare å finne oppslagsord Aasen mente hørte til en mulig standard. I tillegg til uthevede ord på typisk oppslagsplass tok vi bare med uthevede ord med ordklassemarkering. Analysen resulterte i om lag 36 000 artikler og 42 000 oppslagsord. I tillegg kom 10 000 dialektformer som står i den løpende teksten. Den senere manuelle gjennomgangen, har resultert i 7 000 ekstra oppslagsord. Dette skyldes tekstlige og språklige forhold som det er gjort nærmere rede for i vårt andre bidrag (Grønvik, Ore & Minde 2023). Aasens grammatikk foreligger også som digital tekst og er tilgjengelig på nettsidene til Nynorsk kultursentrum, men det er ikke foretatt noen videre oppmerking eller analyse av teksten.

Vi bestemte oss for å lage en egen ordbank for Aasen-standarden og holde den separat fra ordbanken for nynorsk. På et senere tidspunkt kan en eventuelt lage en felles ordbank over nynorsk fra 1864 til i dag. Den største forskjellen mellom ordbanken for moderne nynorsk og den for Aasen-normalen finner en i paradigmene. Aasen-standarden opererer med flere morfologiske trekk, og paradigmene har dermed flere linjer, slik figur 2 viser.

P-id	Linje	Omskrivingsregel	Trekk	forklaring	Fullform	P-id	Ordklasse	Utdjuping	Eksempel	Grunngjeving
942	1o+		eint ub	Eintal ubunden	bok	
942	2o+i		eint bu	Eintal bunden	boki					
942	3ø+er		fl ub	Fleirtal ubunden	bøker	942	subst fem appell	OmlydV O/Ø	bok	Aasen 1864 §171
942	4ø+erna		fl bu	Fleirtal bunden	bøkerna	
942	5o+enne		dat eint	Dativ eintal	bokenne					
942	6o+om		dat fl	Dativ fleirtal	bokom					
942	7o+ar		gen eint	Genitiv eintal	bokar					
942	8o+a		gen fl	genitiv fl	boka					

FIGUR 2 Et paradigme for feminine substantiv med O/Ø-omlyd

For hvert paradigme i Aasen-ordbanken er det en referanse til den eller de paragrafene i Aasens grammatikk som begrunner paradigmet. Aasen-ordbanken har i alt 412 paradigmer mot den moderne ordbankens 580. Dette skyldes nok at den moderne ordbanken har en rekke spesialparadigmer for enkeltord. Ser vi på antall grunnord, har Aasen-ordbanken om lag 46 000 og den moderne 105 000. Av genererte entydige fullformer er det rundt 480 000 i Aasen-ordbanken og 580 000 i den moderne. Så vi ser at den rikere Aasen-morfologien genererer relativt sett flere ordformer.

4. En første dekningsstest

Som en første test på dekningsgraden ønsket vi å kjøre fullformlistene fra Aasen-ordbanken og den moderne ordbanken mot et sett av tidlige nynorsktekster og valgte en samleutgave av verkene til den norske forfatteren og journalisten A.O. Vinje (1818–1870). Vinje døde tre år før Aasens ordbok (Aasen 1873) kom ut, så han kan ikke ha brukt Aasen 1873. Men de hadde i mange år tett kontakt. En antar at Aasen ga Vinje råd om rettskrivningen han skulle bruke, men han hadde også sine egne meninger om det. De viktigste valgene Vinje gjorde, ble publisert som et lite tillegg i hans tidskriftet *Dølen* (Vinje 1859). Vinje skrev for det meste artikler og essay. På midten av 1800-tallet var det i Norge ennå ikke noen tradisjon for å skrive lengre litterære tekster som romaner, noe som skiller Norge fra land som Frankrike og Storbritannia. Vinjes egne tekster i de samlede verkene utgjør ikke mer enn 590 000 løpende ord når de danske tekstene og kommentarapparatet holdes utenfor.

TABELL 1. Resultatet av å kjøre ordbanklistene mot Vinjes skrifter. Det er mange bøyingshomografer og ingen disambiguering, så det som ikke er funnet, er det mest interessante.

	Alle ordformer (tokens)		Entydige ordformer (types)	
I begge:	460 674	78 %	10 324	25 %
bare hos Aasen:	32 110	5 %	4 254	10 %
bare nynorsk (2012):	31 261	5 %	6 114	15 %
I alt:	524 045	89 %	20 692	51 %
Ikke funnet:	65 561	11 %	20 069	49 %
Ordformer i tekstene i alt	589 606		40 761	

Testen vi utførte, ble satt opp som følger: Fra Aasen-ordbanken brukte vi en liste med 520 000 entydige tripler (ordform + POS + informasjon om morfologiske trekk) som bestod av 290 000 entydige ordformer. En tilsvarende liste med 585 000 tripler ble lagd fra ordbanken for moderne nynorsk. Den utgjorde 416 000 entydige former. Her var utvalget begrenset til de anbefalte formene i gjeldende rettskrivning. Et lite, håndlagd konkordansprogram gikk gjennom teksten, isolerte ordformer og sjekket dem mot listene. Resultatet er vist i tabell 1.

Det totale antallet av entydige ordformer (types) i Vinje-tekstene er om lag 40 000. Av disse forekommer 23 360 en gang, mens 146 forekommer mer enn 500 ganger, noe som er i rimelig samsvar med Zipfs lov (Wikipedia 2023). Av de 146 ordformene med høyest frekvens finnes 128 i begge ordbankene, 8 bare i Aasen-ordbanken, 3 bare i nynorskordbanken, og 7 er Vinje-spesifikke. Av dem som bare forekommer i Aasen-ordbanken, er seks bøyde former som har endret staving etter 1873, og de to siste er grunnformer. De tre som bare forekommer i nynorskordbanken, er bøyde former som er blitt endret etter 1873. Flere av de Vinje-spesifikke formene betraktes i dag som Vinjes signaturformer. Han brukte 'ikki' istedenfor Aasens 'ikkje', og preposisjonen 'af' brukte han både i norske og danske tekster. Forekomstene av 'ikke' skyldes dels at det står noe dansk igjen i teksten vi har analysert, men også vakling hos forfatteren. Andre avvikende former kan skyldes at han hadde andre ortografiske vaner enn Aasen.

Omtrent 51 % av typene er funnet i én av eller begge ordbankene, mens 49 % ikke finnes i noen av dem. Vi har gjort en mer detaljert klassifisering av ordformene på 'a-' (i alt 1 449). Nedenfor har vi kort kommentert disse ordformene:

- 1) Ordformene som er funnet i begge ordbankene, hører til kjernevokabularet i nynorsk (inkludert bøyde former). Antallet sammensetninger er relativt lite. Vi finner få importord siden Aasen oftest ikke tok disse med i ordboka. Det er få tvilstilfeller siden Aasens rettskrivningsnorm er veldig klart definert.
- 2) De ordformene som bare finnes i Aasen-ordbanken, er for det meste bøyde former som ikke lenger brukes i moderne nynorsk, for eksempel regelrette adjektiv som ender på «-ad». Endelsen er nå redusert til «-a».
- 3) Av de ordformene som bare finnes i Nynorsk-ordbanken, er 40 % ikke-germanske importord, 17 % navneformer og om lag 10 % danske ordformer som i dag er blitt en del av nynorsk. De resterende 33 % er former av grunnord som er i samsvar med Aasens rettskrivning, men mangler i ordboka (28 %) eller hører til Vinjes egen og mer heterogene rettskrivning (5 %). Vinje publiserte tekster to ganger i uken, og når han var i tvil, ser det ut til at han valgte ordformer fra hans egen dialekt i Vest-Telemark. Noen av disse har senere blitt tatt med i moderne nynorsk.
- 4) Ideelt sett burde flere av ordformene som ikke finnes i noen av ordbankene, vært analysert. Men ut fra analysen av ordene på «a-» kan vi med stor sikkerhet si at disse «Vinje-ordene» faller i følgende grupper: a) danske ordformer (fra sitater og kortere danske tekster), b) importord der mange nå har fått en fornorsket stavemåte, c) ordformer som er i samsvar med Aasens rettskrivning, men som ikke finnes i Aasens ordbok, d) ord som er særegne for Vinje, for eksempel ord som er ortofont stavet, og som ofte er påvirket av Vinjes egen dialekt, e) ord fra latin og moderne språk, særlig engelsk.

Vi kan konkludere med at Vinje og Aasen var ganske enige om hvordan en norsk rettskrivning skulle være. Blant de frekvente ordformene er det få uoverensstemmelser, men de er iøynefallende i Vinjes tekster siden det nettopp gjelder former som forekommer så ofte. Vinje var en forfatter og

journalist, og det viktigste for ham var å skrive slik at han ble lest. Om nødvendig lagde han nye ord, og i den prosessen virker det som norsk tale-mål var hans språklige kompass.

5. Videre arbeid

Norsk er i likhet med de andre skandinaviske språkene et germansk språk og har et produktivt system for ordsammensetning. Dermed er antallet entydige ord (typer) ubegrenset for alle praktiske formål. En fullformliste kan aldri være uttømmende. En løsning på dette problemet er å bruke en såkalt sammensetningsanalysator, altså et program som markerer skillet mellom mulige sammensetningsledd. For Aasen-prosjektet vil en sammensetningsanalysator være et neste skritt. Vi planlegger å teste en slik analysator som er utviklet for Oslo–Bergen-taggeren, se Hagen Johannesen & Nøklestad (2000) og Nøklestad (2022a, 2022b).

Videre er det nødvendig å utvide grunnordlisten for eldre nynorsk. På lengre sikt bør en også inkludere nyere rettskrivninger fra tidlig 1900-tall. For norsk bokmål fra 1800-tallet finnes det ikke noe klart startpunkt. En ordbank basert på ordforrådet i Molbech (1859) kunne være nyttig og kanskje bli et dansk-norsk samarbeidsprosjekt?

Referanser

- Atkins, Beryl T. Sue & Michael Rundell 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Eng, Jan 2014. IBMs leksikografiske prosjekt for norsk 1984–1991. *Maal og Minne* 106:1, 67–101. <<http://ojs.novus.no/index.php/MOM/article/view/225>>. Hentet september 2022.
- Grønvik, Oddrun 2016. The lexicography of Norwegian. I: Hanks, Patrick & Gilles-Maurice de Schryver (red.), *International Handbook of Modern Lexis and Lexicography*. Berlin, Heidelberg: Springer, 1–34.
- Grønvik, Oddrun, Christian-Emil Smith Ore & Trond Minde 2023. Eit ikon kategorisert gjennom ein fullformgenerator. Om Ivar Aasens Norsk Ordbog med dansk Forklaring (1873). I: Holmer, Louise et al. (red.), *Nordiska studier i lexikografi* 16. Lund & Göteborg: Nordiska föreningen för lexikografi, 67–83.

- Hagen, Kristin, Janne Bondi Johannessen & Anders Nøklestad 2000. A Constraint-Based Tagger for Norwegian. I: Lindberg, Carl-Erik & Steffen Nordahl Lund (red.), *17th Scandinavian Conference of Linguistics. Vol. I*. Odense: Institut for Sprog og Kommunikation, Syddansk Universitet, 31–47.
- Hagen, Kristin & Anders Nøklestad 2010. Bruk av et norsk leksikon til tagging og andre språkteknologiske formål. *LexicoNordica* 17, 55–72. <<https://tidsskrift.dk/lexn/article/view/18624>>. Hentet september 2022.
- Henrichsen, Peter Juel 2023. Det Centrale Ordregister. Et indeks for det danske ordforråd – en gave til dansk sprogteknologi. I: Holmer, Louise et al. (red.), *Nordiska studier i lexikografi* 16. Lund & Göteborg: Nordiska föreningen för lexikografi, 113–126.
- Hovdenak, Marit 2006 [1986]. *Nynorskordboka. Definisjons- og rettskrivingsordbok*. 4. utg. Oslo: Samlaget.
- Koskenniemi, Kimmo 1990. Finite-State Parsing and Disambiguation. I: Karlgren, Hans (red.), *COLING'90. Proceedings of the 13th International Conference on Computational Linguistics. Vol. 2*. Stroudsburg: Association for Computational Linguistics, 229–232.
- Molbech, Christian 1859. *Dansk Ordbog indeholdende det danske Sprogs Stammeord tilligemed afledede og sammensatte Ord, efter den nuværende Sprogbrug forklarede i deres forskellige Betydninger, og ved Talemaader og Exempler oplyste*. København: Gyldendalske Boghandlings Forlag.
- Nøklestad, Anders 2022a. *The Oslo-Bergen Tagger*. <<https://github.com/noklesta/The-Oslo-Bergen-Tagger>>. Hentet september 2022.
- Nøklestad, Anders 2022b. *The Compound analyzer software*. <<https://github.com/textlab/mtag>>. Hentet september 2022.
- Ordbøkene 2023. *Bokmålsordboka og Nynorskordboka*. <<https://ordbokene.no/>>. Hentet mars 2023.
- Ore, Christian-Emil Smith 1998. Metaordboken – et rammeverk for Norsk Ordbok? I: Gellerstam, Martin, Kristinn Jóhannesson, Bo Ralph & Lena Rogström (red.), *Nordiska studier i lexikografi* 5. Göteborg: Nordiska föreningen för lexikografi, 250–270. <<https://tidsskrift.dk/nsil/article/view/19466/17092>>. Hentet september 2022.
- Ore, Christian-Emil Smith 2012. Nettordbøker og Norsk Ordbok – hvordan etablere en vitenskapelig nettordbok. I: Eaker, Birgit, Lenn-

- art Larsson & Anki Mattisson (red.), *Nordiska studier i lexikografi* 11. Oslo: Nordiska föreningen för lexikografi, 488–499. <<https://tidsskrift.dk/nsil/article/view/19363/16988>>. Hentet september 2022.
- Ore, Christian-Emil Smith 2016. Gamle ordbøker og digitale utgaver. I: Gudiksen, Asgerd & Henrik Hovmark (red.), *Nordiske Studier i Leksikografi* 13. København: Nordisk forening for leksikografi, 203–216. <<https://tidsskrift.dk/nsil/article/view/111227/160278>>. Hentet september 2022.
- Ore, Christian-Emil Smith & Oddrun Grønvik 2018a. Bokmål og nynorsk samindeksert – Metaordboka som verktøy for jamføring og utforskning av ordtilfang. I: Svavarsdóttir, Ásta, Halldóra Jónsdóttir, Helga Hilmisdóttir & Þórdís Úlfarsdóttir (red.), *Nordiske Studier i Leksikografi* 14. Reykjavík: Nordisk Forening for Leksikografi, 87–95. <<https://tidsskrift.dk/nsil/article/view/117619>>. Hentet september 2022.
- Ore, Christian-Emil Smith & Oddrun Grønvik 2018b. Comparing Orthographies in Space and Time through Lexicographic Resources. I: Čibej, Jaka, Vojko Gorjanc, Iztok Kosem & Simon Krek (red.), *Proceedings of the XVIII EURALEX International Congress. Lexicography in Global Contexts*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani, 159–172. <<https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/118/211/2923-1>>. Hentet september 2022.
- Ore, Christian-Emil Smith, Oddrun Grønvik & Trond Minde 2022. Word Banks, Dictionaries and Research Results by the Roadside. I: Klosa-Kückelhaus, Annette, Stefan Engelberg, Christine Möhrs & Petra Storjohann (red.), *Proceedings of the XX EURALEX International Congress Dictionaries and Society*. Mannheim: IDS-Verlag, 321–333 <<https://doi.org/10.14618/phpy-6r66>>. Hentet mars 2023.
- Ore, Christian-Emil Smith 2020. Å ta Hans Ross på ordet: Ross' ordbok i relasjon til Aasens med Metaordboka som verktøy. I: Sandström, Caroline, Ulla-Maija Forsberg, Charlotta af Hällström-Reijonen, Maria Lehtonen & Klaas Ruppel (red.), *Nordiska studier i lexikografi* 15. Helsingfors: Nordiska föreningen för lexikografi, 253–262. <<https://tidsskrift.dk/nsil/article/view/124027/170987>>. Hentet september 2022.
- Vinje, Aasmund Olavsson 1916–1921. *Skrifter i Samling I–V*. Kristia-

- nia: J.W. Cappelens forlag.
- Walton, Stephen John 1996. *Ivar Aasens kropp*. Oslo: Samlaget.
- Wangensteen, Boye 2005 [1986]: *Bokmålsordboka. Definisjons- og rettskrivningsordbok*. 3. utg. Oslo: Kunnskapsforlaget.
- Wikipedia 2023. *Zipf's law* <https://en.wikipedia.org/wiki/Zipf%27s_law>. Hentet mars 2023.
- Aasen, Ivar 1848. *Det norske Folkesprogs Grammatik*. Kristiania: Det Kongelige Norske Videnskabers-Selskab. Trykt hos Werner & Comp.
- Aasen, Ivar 1850. *Ordbog over det norske Folkesprog*. Kristiania: Det Kongelige Norske Videnskabers-Selskab. Trykt hos Carl C. Werner & Comp.
- Aasen, Ivar 1864. *Norsk Grammatik*. Christiania: Mallings Bogtrykkeri.
- Aasen, Ivar 1873. *Norsk Ordbog med dansk Forklaring*. Christiania: Mallings Boghandel.
- Aasen, Ivar 2003. *Norsk Ordbog med dansk Forklaring* (Ny utg. ved Kristoffer Kruken og Terje Aarset). Oslo: Samlaget