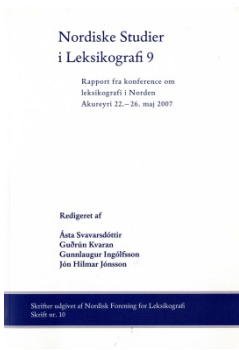


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Norrøn metaordbok – en rapport fra en øvelse	
Forfatter:	Mette Gismerøy Ekker og Christian-Emil Ore	
Kilde:	Nordiska Studier i Lexikografi 9, 2008, s. 89-104 Rapport fra Konference om leksikografi i Norden, Akureyri 22.-26. maj 2007	
URL:	http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive	

© Nordisk forening for leksikografi

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“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Søgbarhed

Artiklerne i de ældre Nordiske studier i leksikografi (1-5) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

Norrøn metaordbok – en rapport fra en øvelse

The term Old Norse usually denotes the (written) languages used in Iceland until 1540 and in Norway until 1370. Old Norse was a weakly normalised language with a wide variety of forms. The scholarly study of Old Norse has long traditions and several dictionaries have been published. *The Dictionary of Old Norse Prose* (ONP) is a current major initiative but so far covering only letters A-E. J.Fritzner's dictionary, FRI, (1883–1896,1974) and F.Jónsson's *Lexicon Poeticum*, LP, (1931) still remain the major dictionaries. The dictionary group of the Medieval Nordic Text Archive (Menota) has as one of its main objectives to establish a framework for linking the disparate Old Norse lexicographic resources. The fragmented state corresponds closely to the state of the *New Norwegian* lexicography 10 years ago. Thus it was a natural choice to try to apply the methods and tools to Old Norse developed for the *Norwegian Dictionary 2014* project (NO2014), the so called Meta Dictionary, a tool for systematizing lexicographical material for weakly normalised languages. In its basic form the Meta Dictionary is an electronic form thesaurus linking realizations of the lemmas. In this pilot project we have focused on the letter B and merged the headword lists of ONP, FRI, LP and a word list based on five lemmatised texts and 200 lemmatised charters. The process and the structure of the resulting Meta Dictionary are divided in three levels of normalisation. All the original forms are kept and made searchable.

1. *Normalisation*: The first step towards merging was simply to ensure one common system for normalised spelling.
2. *Citation form*: The wordlists follow different principles in choice of the form of the cited headword, e.g. nom.sg versus nom.pl. Consequently we had to make sure that all the wordlists use the same forms.
3. *Meta level, further merging of related word forms*: The last step is to connect the multitude of related word forms, both between and within lists, in meta lemmas. Such related forms can be nouns differing in gender, or small differences in word formation.

We also briefly discuss the linking of the medieval material to the modern dictionary projects, like NO2014. We conclude that such a linking at least for Norwegian will involve considerable manual processing. The realization of an Old Norse Meta Dictionary should also rely on a close cooperation with the ONP project.

1 Bakgrunn

Norrønt er det språket vi finner i Norge og på Island i middelalderen. Det nedfeller seg i skriftlige kilder fra rundt 1150. I Norge pleier en å regne den norrøne periode til ca. 1370 og på Island til første halvdel av 1500-tallet. Språket var ikke statisk i denne perioden. Der var dialektale forskjeller og det skjedde endringer over tid. Det overleverte materialet viser både ortografisk og språklig variasjon. Norrøn og nordisk filologi har utviklet seg som disiplin gjennom nærmere 200 år. Tekstene i utgaver og ordbøker for nordiske middelalderspråk preges av et stort mangfold med hensyn på normering, antatt leksikalisering og ortografi. Slik sett er det en dobbel varians for norrønt, originaltekstene er svakt normert og metadisiplinen har utviklet seg. Situasjonen er dermed ikke ulik den vi finner for nynorsk skriftradisjon og leksikografi.

De to store pågående ordbokprosjektene for norrønt og nynorsk, *Ordbog over det norrøne prosasprog* (ONP) og *Norsk Ordbok* (NO), har også mange likhetstrekk. Begge ble grunnlagt på 1930-tallet, og for begge prosjektene var det på slutten av 1990-tallet åpenbart at den daværende redaksjonskapasiteten ville kreve ytterligere flere tiårs redigering før verkene var komplette. For begge prosjektenes vedkommende er det i de senere år blitt foretatt en omfattende digitalisering av kildematerialet. NO har dessuten blitt fullfinansiert med en planlagt ferdigstilling i 2014.

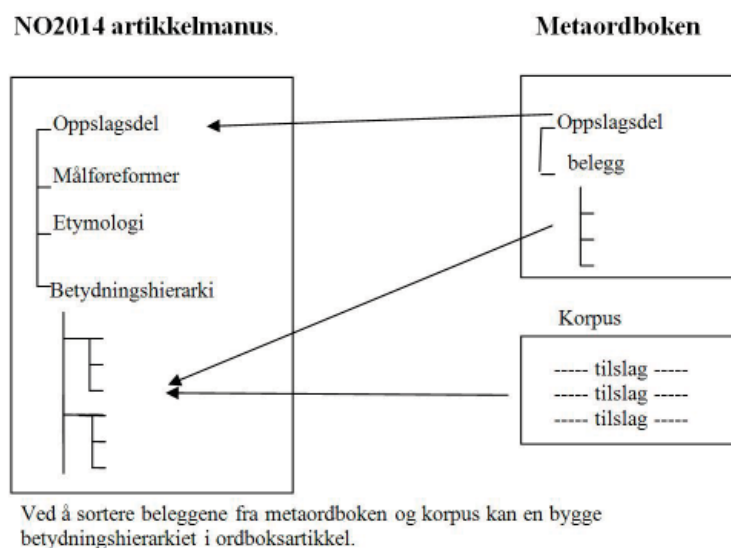
Selv om situasjonen for norrøn leksikografi minner om situasjonen for den nynorske leksikografien på slutten av 1990-tallet, er rammebetingelsene noe annerledes og det organisatoriske atskillig mer komplekst. Norrønt er ikke et levende språk og heller ikke bare knyttet til ett land. Samtidig som norrønt og norrøn tekstutgivelse også er et internasjonalt forskningsfelt, er språket en vesentlig del av Islands og Norges kulturarv.

ONP ble startet på 1930-tallet da Island var underlagt Danmark og må sann sett sies å være et dansk prosjekt. Norge har siden 1940-tallet hatt et eget Gammelnorsk Ordboksverk (GNO) med det formål å dokumentere alle sider av det gammelnorske språket. GNO har mellom annet utgitt et supplementsbind til J. Fritznors ordbok og arbeider med å få realisert *Mellomnorsk ordbok* som skal beskrive norsk i den perioden ONP ikke dekker, 1370–1580. De leksikografiske hovedverkene for norrønt er fortsatt J. Fritznors *Ordbog over det gamle Norske sprog* (1883–1896, 1972) og F. Jónssons *Lexicon Poeticum* (1931). I tillegg kommer de tre første bindene av ONP. Hele verket vil neppe bli ferdig redigert på flere tiår. Det finnes altså ingen fullstendig og samordnet leksikografisk beskrivelse av norrønt.

I løpet av de siste 15 årene har det likevel skjedd en rekke positive ting. Det har vært en tiltagende produksjon av elektroniske utgaver av nordiske middelalderhåndskrifter, og standardordbøkene er blitt gjort elektronisk tilgjengelig. Ved ONP har en også digitalisert store deler av grunnlagsmaterialet. I 2001 ble det

nordiske nettverkssamarbeidet Medieval Nordic Text Archive (Menota, www.menota.org) etablert. Menota er et nettverk av nordiske forskningsinstitusjoner og -selskaper som har til hensikt å bevare, formidle og utgi middelaldertekster i elektronisk form. Menotahåndboken har gitt oss en utmerket standard for koding av nordiske middelalderhåndskrifter. Samarbeidet om publisering av elektroniske nordiske middelaldertekster har kommet i godt gjenge. Det har kommet på plass et elektronisk tekstarkiv med et tilhørende lemmatiseringsverktøy og et rammeverk for et elektronisk korpus. Menotas gruppe for leksikografi, Nordiske middelalderordbøker, (www.edd.uio.no/menota) ser på mulig samarbeid og deling av leksikalske ressurser mellom de ulike ordboksprosjektene for middelalderspråk i Norden. Arbeidet som er beskrevet i denne artikkelen er gjort i den forbindelse. Vi har i vår øvelse sett på hvordan en kan knytte sammen de leksikalske ressursene i et felles rammeverk for søk i de ulike digitale ordboksressursene og tekstkorpusene.

2 Hva er norrøn metaordbok?



Figur 1: Norsk Ordboks redigeringssystem, Metaordboken og korpus

Middelalderspråk som norrønt er svakt normerte språk med et stort utvalg av ulike ordformer. For å optimalisere søk og analyser bør en størst mulig del av ordene i de løpende tekstene være lemmatisert og påført bøyninginformasjon. Det finnes ingen utviklet morfologisk analysator (tagger) for norrønt. Som et viktig ledd i å sikre god funksjonalitet trenger korpuset å ha en formtesaurus koblet mot eksisterende ordbøker for norrønt. Denne formtesauren eller metaordboken vil være ryggraden i sammenkoblingen av de leksikalske enkeltdatabasene. Den vil også kunne danne forbindelsen til de moderne ordboksverkene for islandsk og

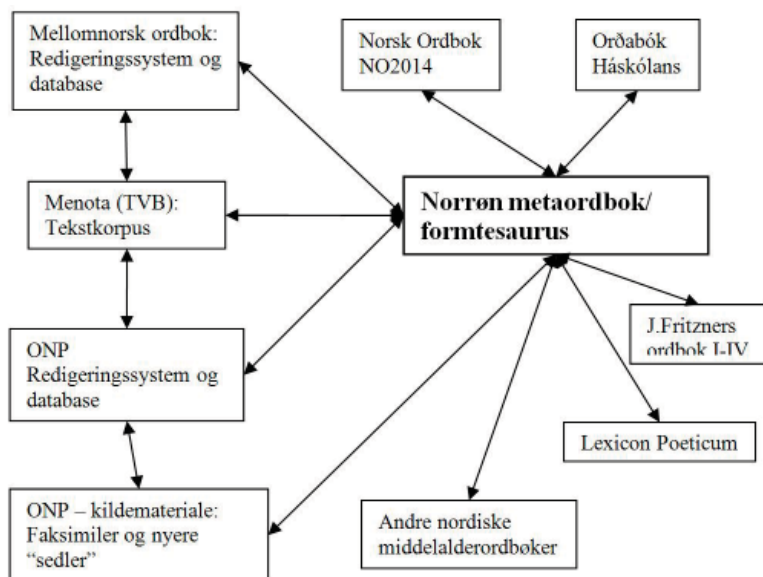
norsk, *Orðabók Háskólans* og *Norsk Ordbok*, som beskriver språkene fra etterreformatorisk tid til i dag. Vi mener at et slikt rammeverk kan bygges over samme lest som Metaordboksystemet for Norsk Ordbok. Metaordboken er et velprøvd verktøy som har blitt brukt av NO2014 de siste 6 årene. Det er kanskje ikke så velkjent, så det er naturlig å bruke litt plass på å forklare bakgrunnen og hva det går ut på.

Nynorsk leksikografi var på 1990-tallet i samme situasjon som norrøn leksikografi med det unntak at norrøn leksikografi er spredt over flere land og prosjekter. I den nettbaserte virkeligheten spiller det siste en mer organisatorisk enn praktisk rolle. Som et ledd i revitaliseringen av de store norske ordboksprosjektene gjennomførte en på 1990-tallet en storstilt digitalisering av tekstsamlinger, kortarkiv og ordbøker, og det ble lagd systemer for morfologisk merking av løpende tekst. Dette arbeidet har siden fortsatt. Det revitaliserte Norsk Ordbok prosjektet, NO2014, har nå også et tekstkorpus på mer enn 30 millioner ord. I løpet av siste halvpart av 1990-tallet ble det utviklet en modell for hvordan disse ressursene, både det gamle materialet og det nye elektroniske tekstkorpuset, kunne bindes sammen i et rammeverk for leksikografisk arbeid for å understøtte redigeringsarbeidet, men også for å gjøre materialet tilgjengelig for andre forskere og den interesserte allmennheten. Dette rammeverket ble kalt en *metaordbok* og grunnideene er beskrevet i Ore 1999.

Metaordboken er et verktøy for å samle belegg og normere svakt normerte språk. Hver artikkel i Metaordboken er en samling referanser til kilder som beskriver det aktuelle ordet eller gir brukseksempler. Hver artikkel kan ha et eller flere normerte oppslagsord med tilhørende bøyningsinformasjon. Disse oppslagsordene kan være konstruert i henhold til en gitt rettskrivningsnorm. De brukes som bindeledd til de egentlige ordboksartiklene i Norsk Ordbok. For norrønt vil en kunne ha oppslagsord til meta-artiklene både i henhold til GNOs standard og til ONPs.

Fullt utbygd og korrekt anvendt skal Metaordboken til NO2014 avspeile prosessen med å skape ordboksartikler helt fra ubehandlede samlinger, via omstruktureringer og systematiseringer til en eksplisitt beskrivelse av hva de rå samlingsdataene betyr, se figur 1. I prinsippet skal man ha tilgang til alt materiale som er brukt i redigeringsprosessen, også det som ikke er kommet med i ordboksartikelen. På den måten vil man komme et steg videre med å gjøre en ordbok vitenskapelig etterprøvable.

Når vi nå har sett på mulighetene av å lage en metaordbok for norrønt, er det uten intensjoner om å lage noen felles nordisk middelalderordbok. Hensikten er å få til et system for å koble sammen alle elektroniske tekster og leksikalske ressurser. Den norrøne metaordboken er ment som et hjelpemiddel både for de norrøne ordboksprosjektene og for tekstdelen av Menota-samarbeidet, se figur 2. Den vil også være en måte for elektroniske versjoner av de to store ordbøkene for nyere tid, *Orðabók Háskólans* og *Norsk Ordbok*, å hekte seg på til det førreformatoriske materialet.



Figur 2: Noen mulige forbindelser gjennom en norrøn metaordbok

3 Grunnlagsmaterialet

Metaordboken kan i sin enkleste form sees som en samindeksering av en rekke kilder og belegg. Vi har foreløpig begrenset oss med visse unntak til denne enkle formen og kun sett på ord som begynner på bokstaven *B* for å finne ut mer om arbeidsmengde og problemer. Kildene vi ønsker å koble er hovedordbøkene for norrønt. Det vil si ONP, J. Fritzners ordbok og F. Jónssons *Lexicon Poeticum*. For å få til en kobling med korpusmaterialet har vi også tatt med en liste over ordforrådet i fem større tekster¹ og 200 norskspråklige diplomer fra før 1308. De elektroniske tekstene er i xml-format og lemmatiseringen er basert på GNOs ordseddelsamlinger. Tekstene svarer til om lag 420 000 løpende ord.

Tabellen i figur 3 gir en oversikt over forholdet mellom løpende ord (tokens), forskjellige ordformer (types på formnivå) og grunnord (types) på lemmanivå i de fem tekstene og diplomene. I snitt utgjør lemma på bokstaven *B* om lag 6,2 prosent av hele lemmautvalget, mens på token-nivå ligger *B*-ordformer nede på 3,3 prosent. Det skyldes antagelig at det ikke finnes så mange høyfrekvente funksjonsord på bokstaven *B*. Men det er verdt å legge merke til at det relative antallet oppslagsord på bokstaven *B* i de tre ordbøkene er i godt samsvar med den faktiske fordelingen i tekstutvalget. For ONP må spesielt totaltallene tas med en klype salt. De er hentet fra nettordlisten. Tallene i raden markert med "hovedord" er

¹ Konungs skuggsjá, AM 243 b α fol; Óláfs saga ins helga, Upps DG 8; Strengleikar, Upps DG 4–7; Þiðriks saga af Bern, Holm perg 4 fol; Thómass saga erkibiskups, Holm perg 17 4to.

de ord som ikke er markert som poetisk, navn, fremmedord, eller yngre islandsk eller norsk. De aller fleste oppslagsordene i *Lexicon Poeticum* er derfor ikke med i ONPs hovedordliste slik den fremstår på nettet. Tilsvarende er ikke navn eller latinske ord i tekstene med der. Yngre norske og islandske ord i Fritznors ordbok er heller ikke med i ONPs hovedordliste.

Tekst	Forskjellige grunnord			Forskjellige ordformer			Løpende ordformer		
	B	alle	%	B	alle	%	B	alle	%
Thóm.	186	3 709	5,0	699	14 284	4,9	2 430	64 380	3,8
Kgs.	280	4 116	6,8	951	14 194	6,7	1 780	56 226	3,2
Þiðr.	235	4 210	5,6	1 305	24 334	5,4	4 139	147 766	2,8
Str.	147	2 597	5,7	487	9 889	4,9	1 100	42 085	2,6
Ólaf.	183	3 234	5,7	603	9 738	6,2	1 218	41 060	3,0
Dipl.	312	4 806	6,5	1 235	20 111	6,1	3 050	66 175	4,6
Alle	841	13 598	6,2	4 544	75 567	6,0	13 717	417 692	3,3

Ordbok	B	Hele	%
Fritzner, artikler	2 400	38 500	6,2 %
Lexicon Poeticum, artikler	1 150	1 8000	6,4 %
ONP, ordliste, hovedord	3 900	72 000	5,4 %
ONP, ordliste, alle ord (antall artikler a-em)	5 600	91 700	6,1 %

Figur 3: Lemma- og ordformfrekvenser i de ulike tekstene og i de to hovedordbøkene for norrønt.

4 En tre-trinns prosess for samkjøring av ordlister

Et konkret mål for vår øvelse var å samkjøre de fire ordlistene slik at like lemma i hver av listene ble koblet sammen i en formtesaurus. For å gjennomføre dette satte vi opp følgende tre-trinns prosess:

1. felles normalisering
2. valg av oppslagsform og grammatisk klassifisering
3. metanivå, en videre sammenslåing av relaterte ordformer

4.1 Trinn 1: normalisering

Normalisering innenfor norrønt er konstruert i nyere tid etter mønster av islandsk ortografi. Det finnes flere normaliseringsvarianter. Ingen av de fire ordlistene våre følger nøyaktig samme prinsipper for normalisering. En oversikt over vesentlige forskjeller gis i figur 4.

Det kan være snakk om rent ortografiske ulikheter, som at ONP gjengir lang 'ø' med aksent for markering av vokallengde, mens de andre listene bruker 'œ'-ligatur. Ulik normaliseringspraksis kan også gjenspeile geografisk variasjon, som

når GNOs ordformer viser gammelnorske trekk ved at 'h' ikke står foran 'l', 'n' og 'r' i framlyd.

Fenomen	ONP	GNO	Fritzner	LP
æ	ǣ	æ	æ	æ
langø	ø	œ	œ	œ
kort ø	ø	ø	ö	ø
o med kvist	o	o	ö	o
lango med kvist	á	á	á	ȝ
øy/ey	ey	øy	ey	ey
konsonantisk i	j	i	j	j
pt/ft	ft	pt / (ft)	pt / (ft)	pt
langK + ny K	ggð, lld	gð, ld	gð, ld	ggð, ld
hl, hn, hr	hl, hn, hr	l, n, r	hl, hn, hr	hl, hn, hr

Figur 4: Normaliseringsforskjeller mellom ordlistene

I tråd med Menotas prinsipper valgte vi å følge ONPs standard for normalisering. Mye av samnormaliseringen kunne utføres ved kontrollerte søk-erstatt rutiner. Men enkelte ulikheter mellom listene måtte lukes ut manuelt. Det siste gjaldt f.eks. for ordlistenes ulike realisering av vokalen 'o med kvist'. Dette var en lav, bakre, rundet vokal som oppstod ved overgangen fra urnordisk til norrønt. Den ble ikke værende i språket. I islandsk falt den etterhvert sammen med kort 'ø' og i norsk hovedsakelig med kort 'o'. ONP og LP bruker et entydig spesialtegn for 'o med kvist'. Fritzner følger nyere islandsk ortografi med bokstaven <ö>, mens GNO bruker bokstaven <o>. Fritzner har dermed sammenfall mellom etymologisk kort 'ø' og kort 'o med kvist', mens GNO har sammenfall mellom etymologisk kort 'o' og kort 'o med kvist'. Denne typen variasjon i listene måtte lukes ut manuelt og føres over til ONPs standard.

Til hjelp i arbeidet med å identifisere avvikende ordformer, foretok vi jevnlig flettinger av ordlistene og kjørte ut *diskrepanslister* over de ordformene i Fritzner, GNO og LP som ikke falt sammen med noen ordform i ONP. Ved fletting fulgte vi følgende prinsipper: GNO-listen er ikke homografseparert, så vi skiller ikke mellom homografer av samme ordklasse. I ordlistene er forledd og etterledd egne innførslers, så vi har valgt å skille mellom oppslagsord markert som selvstendige ord og oppslagsord markert som for- eller etterledd. Former som *bakki* m. og *-bakki* m. ble dermed ikke koblet sammen.

Ved samkjøring av ordformer beholdt vi de ulike listenes opprinnelige oppslagsform. For hvert trinn i prosessen, ble det i hver liste lagt til et ekstra felt for lemma og grammatisk kategori, og alle endringer på trinnet ble foretatt i de nye feltene. Figur 5 viser trinn-1 felt etter samnormalisering av ordet *banahogg*. Listenes opprinnelige form står i feltene til høyre.²

² For feltet 'Gram-1' i tabellen i figur 5 har vi ikke fulgt ONPs notasjon. Vi vil etterhvert følge Menotas grammatiske notasjon som baserer seg på anbefalinger fra Eagles (<http://www.ilc.cnr.it/EAGLES96/annotate/annotate.html>). I tabellen er det brukt en midlertidig arbeidsnotasjon.

Liste_ID	Trinn-1	Gram-1	Oppslagsord	Gram
GNO	banahøgg	n	banahøgg	n
ONP	banahøgg	n	banahøgg	sb. n
Fritzner	banahøgg	n	banahögg	n

Figur 5: Trinn-1 felt for banahøgg med 'o med kvist' ført inn.

Etter fullført samnormalisering, trinn 1, kjørte vi ut nye diskrepanslister og regnet ut avviksprosent for hver av listene i forhold til ONP. Av Fritzners totalt ca. 2400 ord på B, kom det 180 ord på diskrepanslisten, hvilket tilsvarer avvik på 7,5 %. For LP og GNO var avviket på henholdsvis 5 % og 16 %. Vi kom dermed ganske langt bare ved å samnormalisere listene.

4.2 Trinn 2: valg av oppslagsform og grammatisk klassifisering

4.2.1 Ulike prinsipper for valg av oppslagsform

De ordene vi fikk ut på diskrepanslisten etter trinn 1, falt i en del klart definerbare kategorier som igjen har sammenheng med ulike prinsipper for valg av oppslagsform. De mest frekvente av disse er vist i figur 6. Eksemplene her er hentet fra Fritzners avvik i forhold til ONP, men de samme kategoriene går igjen også i de andre listene.

Fritzner		ONP		
Substantiv nom. pl.	<i>barir</i> <i>bróðrasynir</i>	Substantiv nom. sg.	<i>bör</i> <i>bróðrasonr</i>	'(lik)bære' 'brødresønner'
Verb refleksiv	<i>bakverpast</i> <i>baðast</i>	Verb infinitiv	<i>bakverpa</i> <i>baða</i>	'forkaste - vende ryggen til' 'bade'
Adverb	<i>berbakt</i> <i>bátlaust</i>	Adjektiv	<i>berbakr</i> <i>bátlauss</i>	'med bar rygg' 'uten båt'
Adjektiv (=perf.part.)	<i>borinn</i>	Verb infinitiv	<i>bera</i>	'bære'

Figur 6: Ulike prinsipper for valg av oppslagsform

Fritzner har ført opp flertallsformen av substantiv som han i kildematerialet utelukkende har funnet belagt i flertall, mens ONP har ført opp entallsformen. Enten har ONP funnet belegg for entallsformen, hvilket vil framgå av ordboksartikkelen, eller også er entallsformen konstruert. Fritzner har likeledes ført opp refleksivform av verb han ikke har funnet belagt i aktiv, mens ONP har ført opp infinitiven av verbet, enten de har funnet belegg for bruk i aktiv eller ikke. De to verbene i figur 6 er forøvrig begge funnet belagt i aktiv. Tilsvarende har Fritzner ført opp adverb der han ikke har adjektivformen belagt, mens ONP har ført opp adjektivet, enten det er en reell eller konstruert form.

Det er klart det kan diskuteres hvorvidt man skal konstruere oppslagsformer

som kanskje aldri har eksistert. Men vi hadde som overordnet prinsipp å følge ONP, og da ble det å identifisere denne typen diskrepans og føre den over til ONPs valg av oppslagsform, en relativt enkel oppgave.

4.2.2 Adjektiviske partisipper

Et eksempel på en mer problematisk samkjøring av ordformer på dette trinnet, er adjektiviske partisipper. I ordlistene finner vi blant annet følgende oppslagsformer med utgangspunkt i verbet 'bære':

Fritzner:	<i>bera</i> v.	<i>borinn</i> adj.	<i>arfborinn</i> adj.
ONP:	<i>bera</i> v.	<i>-borinn</i> adj.	<i>arfborinn</i> adj.

Hos Fritzner er slike partisippformer ført opp som selvstendige lemma og klassifisert som adjektiv. I ONP er de samme partisippformene behandlet under artikkelen til verbet. Ut fra vårt prinsipp om å følge ONPs valg av oppslagsform, burde vi føre Fritzners adjektivpartisipper til verbets infinitiv, dvs. *borinn* adj. føres til lemma *bera* v. Men dersom partisippformen inngår i en fast sammensetning i norrønt, har både ONP og Fritzner klassifisert sammensetningen som adjektiv, f.eks. *arfborinn* adj, 'arveberettiget'. I ONP er etterleddet i tillegg ført opp som adjektiv, med henvisning til verbet. Siden vi har valgt å skille mellom selvstendige ord og ledd, førte dette til følgende to lemma på trinn 2, med tilknyttede ordformer:

Lemma 1, <i>bera</i> v.:	<i>bera</i> v. (ONP)
	<i>bera</i> v. (Fri)
	<i>borinn</i> adj. (Fri)
Lemma 2, <i>-borinn</i> adj.:	<i>-borinn</i> adj. (ONP)

En kompliserende faktor her er at Fritzner ikke nødvendigvis opplyser at en oppslagsform utelukkende opptrer som del av en sammensetning. Vi kan derfor gå ut fra at flere av de adjektiviske partisippene i Fritzner i realiteten er ledd og dermed identisk med adjektivleddene i ONP av typen *-borinn* adj. Men det vil neppe gjelde alle partisippformene i Fritzner, og det virket lite gunstig å behandle adjektivpartisippene ulikt på dette stadiet i prosessen. Derfor valgte vi på trinn 2 å føre alle slike partisipper hos Fritzner til verbets infinitiv.

4.2.3 Oppsummering trinn 2

Diskrepanslistene vi kjørte ut avslutningsvis på trinn 2, viste et ganske oppløftende resultat, særlig for Fritzners vedkommende. Det var nå bare 0,6 % av ordformene i Fritzner som vi ikke hadde klart å føre over til en form i ONP. GNO hadde en noe høyere avviksprosent på 12 %. Dette skyldes hovedsakelig at GNO-listen inneholder mange navn og navneledd, og slike former er ikke tatt med i ONP.

4.3 Trinn 3, en videre sammenslåing av relaterte ordformer

4.3.1 Relaterte ordformer

Etter fullført trinn 2 har vi etablert en formtesaurus som tilsvarer ONPs lemmaliste pluss særformer fra de øvrige listene. ONP er den mest omfangsrike av de norrøne ordlistene, blant annet fordi den har med langt flere variantformer i form av nært relaterte ordformer der de andre listene økonomiserer ordutvalget, se figur 7. En slik ordrikdom er positiv sett fra en bruker av ordbokens side. Når man skal slå opp et ord man støter på i en norrøn tekst, vil man stort sett finne det i ONP. Men denne overflod av nært relaterte ordformer er ikke nødvendigvis like gunstig i en metaordbok, f.eks. med tanke på søk i korpus, eller generering av statistikker over ordforekomster. Dersom man søker etter *barnafaðir*, kan det være av interesse også å få informasjon om relaterte former som *barnfaðir* og *barnsfaðir*. Den ene formen bør ikke utelukke de to andre. Det fører oss til trinn 3 i samkjøringsprosessen, hvor vi går ut over ONPs lemmaliste og slår sammen nært relaterte ord i metalemma, som de tre sammensetningene for 'barnefar' i figur 7. Disse sammensetningene er behandlet i separate oppslag i ONP. Fritzner har kun formen med nullfuge, *barnfaðir*. Ordet forekommer ikke i GNOs materiale eller i LP.

Fenomen	ONP	Fritzner
Substantiv, sterk og svak form	<i>bjartleikr</i> sterk m. <i>bjartleiki</i> svak m. <i>bjarnfell</i> sterk n. <i>bjarnfelli</i> svak n.	<i>bjartleikr</i> sterk m. = 'lys, glans' <i>bjarnfell</i> sterk n. = 'bjørneskinn'
Substantiv med variasjon i grammatisk kjønn	<i>bið</i> f. <i>bið</i> n.	<i>bið</i> n.pl. = 'venting, ventetid'
Svake adjektiv på -a og -i	<i>blóðrisa</i> <i>blóðrisi</i>	<i>Blóðrisa</i> = 'blodsprenget'
Sammensetninger med morfologisk variasjon i 1. ledd	<i>barnfaðir</i> m. <i>barnafaðir</i> m. <i>barnsfaðir</i> m.	<i>barnfaðir</i> m. = 'barnefar'

Figur 7: Eksempler på typer av variasjon i ONP utelatt hos Fritzner

Vi trengte imidlertid kjøreregler for en slik grovere sammenslåing av ordformer. Vi var i den heldige situasjon for bokstaven B at ONP allerede har behandlet den. I ONPs artikler henvises det alltid mellom beslektede ordformer. ONP er dessuten digitalisert og gjort tilgjengelig via internett. Vi kunne derfor hente disse krysshenviisningene fra nettutgaven og generere lister over relaterte former. Men siden ONP ikke skiller mellom henvisninger som angir semantisk slektskap og de som angir formvarians, måtte det en manuell gjennomgang til.

Valg av metalemma har vært gjenstand for en del diskusjon, og vi har til dels også eksperimentert med mulige tilnæringsmåter. Vi kunne velge den etymologisk mest korrekte formen, eller den mest frekvente formen, eller en bestemt fuge i sammensetninger. Men disse framgangsmåtene var ofte svært tidkrevende og passet som regel aldri for alle typer av variasjon mellom relaterte ordformer. I dette prøveprosjektet var det også et poeng å få dataene samkjørt relativt raskt slik at vi kunne få et overblikk over materialet. Derfor valgte vi en enklere og kanskje noe uortodoks metode: Metalemma er den formen som kommer først alfabetisk. Dette er en metode som lar seg klart definere, den er etterprøvable og kan utøves på alle ordformene. Det må også understrekes at noe av ideen i metaordboken er at artiklene kan gis flere metaoppslag, det vil si kan indekseres etter flere ordlister.

4.3.2 Sammensetninger med variasjon i fugen

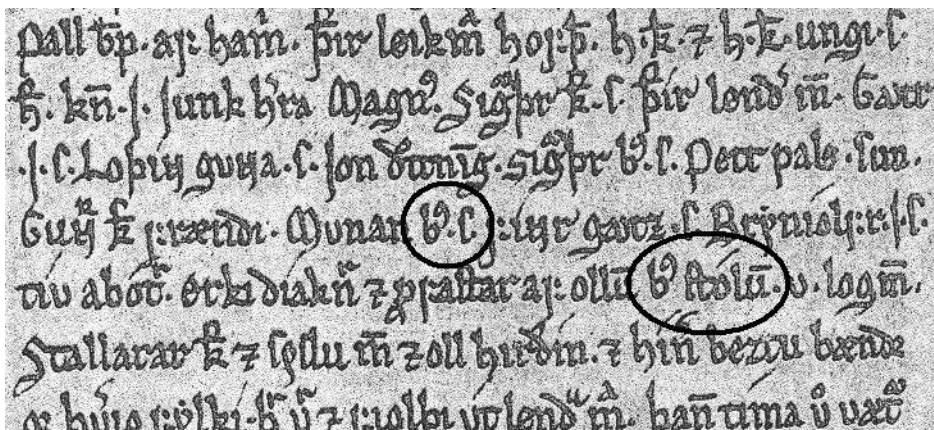
For sammensetningers vedkommende vil den alfabetiske tilnæringsmåten føre til at form med g.pl. i førsteleddet, såkalt a-fuge, dersom den er belagt, blir valgt som metalemma:

metalemma:	<i>barnafaðir</i>		
relaterte former:	<i>barnafaðir,</i>	<i>barnfaðir,</i>	<i>barnsfaðir</i>

I en ferdig metaordbok vil ordlistenes opprinnelige ordformer, samt alle samkjøringsfeltene fra trinn 1 og trinn 2, være søkbare, ikke bare det normaliserte metalemma. Det innebærer at et søk på en form som *barnfaðir*, med nullfuge, vil gi direkte treff på Fritznere og ONPs artikler med denne formen, men også gi tilgang til de to alternative formene i ONP, med a-fuge og s-fuge.

Det faktum at Fritznere i stor grad slår sammen denne typen ordformer, er et argument for at det er forsvarlig også for oss å gjøre det i metaordboken. Et annet argument i samme retning er at ordbøkene, særlig de eldre, dels ekserperer sine ordbelegg fra trykte tekstutgaver av håndskriftene. I middelalderen praktiserte man en utstrakt bruk av forkortede skrivemåter. I trykte utgaver av håndskriftene vil utgiver ofte løse opp forkortelsene. I nyere diplomatariske, dvs. bokstaverette, tekstutgaver er slike oppløste forkortelser som regel tydelig markert. I eldre diplomatariske tekstutgaver er ikke dette nødvendigvis tilfelle, og man kan som bruker av teksten ikke vite hva som er originalens skrivemåte og hva som står for utgivers regning. Som et eksempel kan vi se på følgende skrivemåter for *byskup(s)son* og *byskup(s)stólum* i håndskriftet *Frisbók*³ fra begynnelsen av 1300-tallet, figur nr. 8.

³ AM 45 fol, Codex Frisianus, ca 1300–1325, her gjengitt etter Halldór Hermannsson 1932.



Figur 8: Frisbók s. 123v spalte 452l. 12 *byskvp(s) son* og l. 13 *byskvp(s) stolum*. Understreking markerer oppløste forkortelser.

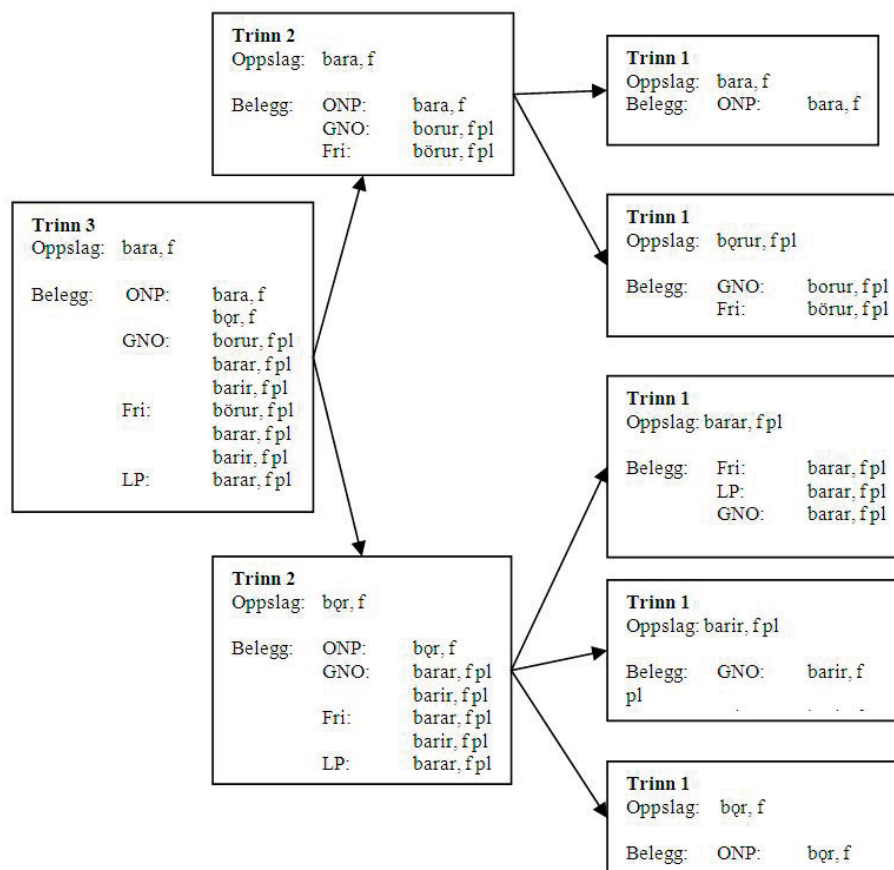
Teksten er her såpass forkortet at vi ikke ut fra skrivemåten kan se hvilken fuge skriveren mon hadde i tankene. I tekstutgaven av Frisbók, Unger 1871, er disse forkortelsene oppløst uten markering som henholdsvis *byskvpsson*, med s-fuge, og *byskvpstolum*, med nullfuge. Utgiver følger det han oppfatter som håndskriftets norm ved oppløsning av forkortelser. ONP er med alle sine ordformer en deskriptiv ordbok, mens Fritzner kan sies å være en mer normativ ordbok. Men grunnlaget for beleggene i ONPs deskriptive samling av ordformer er i enkelte tilfeller allerede normert i og med utgivers valg av oppløsning. Metaordboken viderefører denne normeringstendensen i den grovere sammenslåingen av ordformer i trinn 3.

4.3.3 Metalemma og relaterte former; et eksempel

I det norrøne kildematerialet finner vi flere relaterte ordformer for det som tilsvarer vårt '(lik)bære'. Begrepet forekommer hovedsakelig i flertall, med tre belagte flertallsformer, *barar*, *barir* og *børur*. Disse finner vi som tre separate innførsler i GNO og Fritzner, mens LP bare har formen *barar*. I ONP er de to første formene, *barar* og *barir*, begge ført til en konstruert entallsform *bør f*, mens den siste flertallsformen, *børur*, er ført til en belagt entallsform *bara f*. Dette viser at ONPs tendens til å ta med et stort antall relaterte ordformer til en viss grad dempes noe av ordbokens strengere regler for valg av oppslagsform.

ONPs to oppslagsord er en svak og en sterk form av et feminint substantiv. Det er flere paradigmer knyttet til den sterke formen (nom. pl. *barir* og *barar*), og det er også sammenfall i enkelte bøyningsformer mellom sterk og svak form (f.eks. dat.pl. *børur*.) ONP krysshensier mellom de to oppslagsordene. På et metanivå kan vi se ONPs to oppslagsord som varianter av samme lemma, med ulike paradigmer knyttet til seg. På trinn 3 blir alle de ni innførslene for 'bære' ført sammen i én metaordboksartikkel, se figur 9. I metaordboken vil hver ordform

på hvert trinn i prosessen være søkbar og gi tilgang til alle formene tilknyttet metaartikkelen.



Figur 9: Metalemma bara f. og relaterte ordformer i listene.

5 Noen betraktninger om linjene fram til de moderne ordbøkene

Den norrøne metaordboken skal være en felles inngang til ordbøkene over islandsk og norsk i norrøn tid ved å gruppere sammen ulike skriftrealiseringer av det en ønsker å oppfatte som ord eller lemma. Norrøn metaordbok vil gjennom de ulike ordbøkene koble sammen ordbelegg over en periode på rundt 500 år. Til sammenligning dekker *Norsk Ordbok* norske dialekter i drøye 300 år og *Orða-bók Háskólans* islandsk i 460 år. Det er lange spenn, men ikke oppsiktsvekkende lange. *Oxford English Dictionary* dekker engelsk over en 1000 års periode.

For de norske historisk-orienterte leksikografiske miljøene er en norrøn metaordbok interessant fordi den vil danne grunnlag for arbeidet med en Mellomnorsk ordbok for perioden 1370–1580 og kunne brukes til å lage en forbindelse

mellom *Norsk Ordboks* beskrivelse av tidlig-moderne og moderne norsk. Det er i så fall et spørsmål hvordan en slik forbindelse skal etableres. I de etablerte norske ordbøkene er det relativt få ord som er markert med gammelnorsk eller gammelislandsk opphav. A. Torps nynorske etymologiske ordbok behandler om lag 15 000 stort sett usammensatte ord. Bare 5 000 av dem har henvisning til norrønt eller gammelislandsk. For bokstaven B har de tre nynorske ordbøkene, *Norsk Ordbok*, A. Torps etymologiske ordbok og *Nynorskordboka*, tilsammen drøye 900 ord med slik henvisning. *Norsk Ordbok* har 18 400 oppslagsord på bokstaven B, mens håndordboken *Nynorskordboka* har 7200. Det er pussig om bare fem prosent av nynorske ord skulle ha norrønt opphav. Motsatt vei er det maksimalt 20 prosent av oppslagsordene i ONP og maksimalt 40 % av oppslagsordene i Fritznert som er referert fra de moderne norske ordbøkene. Dette er kanskje mer naturlig, men er verd et nærmere studium. Denne raske kvantitative testen sier oss at en etter all sannsynlighet må utføre et omfattende leksikografisk arbeid for å få på plass en ordentlig kobling mellom *Norsk Ordbok* og det norrøne materialet. Det foreligger ikke nok ordhistorisk informasjon i de eksisterende ordbøkene til å gjøre en særlig god mekanisk kobling.

6 Oppsummering og videre arbeid

Både den alfabetiske tilnæringsmåten og bruk av krysshenvisingene fra ONP, er nyttige hjelpemidler i arbeidet med å føre sammen relaterte ordformer på trinn 3, men de er ikke absolutte prinsipper. De genererte listene over krysshenvisinger bør derfor brukes til manuell gjennomgang av de i ONP relaterte formene.

Den alfabetisk utvalgte metalemmalisten kan tilsvarende sees som et utgangspunkt for videre bearbeiding. I de digitaliserte ordlistene er hovedform og sideform splittet i separate innførsler for å muliggjøre søk på alle former. Ved femininer på *-an* og *-un* (eks: *skipan*, *skipun*) fører ONP formen med u-vokalisme i endelsen som hovedform, mens den alfabetiske tilnæringsmåten vil velge ut formen med a-vokalisme. En manuell gjennomgang av en alfabetisk generert metalemmaliste er derfor både ønskelig og nødvendig. Men i og med at alle ordformer på alle trinn i samkjøringsprosessen vil være søkbare, er valg av metalemma mindre viktig enn valg av former knyttet til dette metalemma. Det vesentlige er at både *skipan* og *skipun* knyttes sammen, ikke hva vi velger å kalle enheten på toppen. Gjennomgang og redigering av den alfabetisk genererte metalemmalisten kan derfor gjøres på et senere stadium.

Sammenslåingen i trinn 3 har i vårt prøveprosjekt basert seg på informasjon som finnes i ONPs ferdigredigerte artikler på nettet. Disse finnes så langt ONP er kommet med artikkelredigeringen, altså til midten av bokstaven E. For resten av alfabetet er der ingen krysshenvisinger å hente ut fra nettversjonen. I ONPs database for resten av alfabetet er slike krysshenvisinger bare satt inn sporadisk. Det vil si at her må materialet gjennomgås manuelt for å identifisere tilsvarende

relasjoner mellom ord, noe som vil bli en langt større og mer ressurskrevende jobb enn vi har utført i dette prøveprosjektet. Det vil også være en forutsetning at dette kan gjøres i samarbeid med ONPs redaksjon.

Litteraturliste og nettsteder

- Halldór Hermannsson 1932 (ed. facs.): Codex Frisianus (Sagas of the Kings of Norway): MS. No. 45 Fol. in the Arnamagnæan Collection in the University Library of Copenhagen, *Corpus codicum Islandicorum medii aevi* 4, Copenhagen 1932.
- Jónsson, Finnur 1931: Lexicon Poeticum antiquæ linguæ septentrionalis, Ordbog over det norsk-islandske skjaldesprog oprindelig forfattet av Sveinbjörn Egilsson, S.L.Møllers bogtrykkeri, København, 1931.
- Fritzner, Johan 1886–1896: Ordbog over det gamle norske Sprog I–III, 4. utgave Universitetsforlaget, Oslo, Bergen Tromsø. 1972, (www.dokpro.uio.no, sjekket 270907).
- Hødnebo, Finn 1972: Rettelser og tillegg til Johan Fritzner, Ordbog over det gamle norske Sprog, Universitetsforlaget, 1972, Oslo.
- Menota – Medieval Nordic Text Archive (www.menota.org, sjekket 270907)
- Norrøn metaordbok www.edd.uio.no/menota/metaordbok.html
- Nynorskorboka, tredje utgave, Det norske samlaget, Oslo 2001.
- Norsk Ordbok 1–6, Det norske samlaget, Oslo 1950–2007, (www.no2014.uio.no (sjekket 270907)).
- Ordbog over det norrøne prosasprog, 1–3, København, www.onp.kua.dk (sjekket 270907).*
- Orðabók Háskólans (www.lexis.hi.is/indexny.html, sjekket 270907).
- Ore, Christian-Emil 1999: *Metaordboken – et rammeverk for Norsk ordbok?* Nordiske studier i leksikografi 6, NFL, Göteborg, 1999.
- Torp, Alf 1919: Nynorsk etymologisk ordbok, Kristiania 1919.
- Unger, Carl Richard 1871 (ed.): Codex Frisianus: En Samling af norske Konge-Sagaer, Christiania, 1871.
- Font brukt for norrøne tegn: *Andron Scriptor Web*, designet av Andreas Stötzner, Leipzig, og distribuert av Medieval Unicode Font Initiative (Mufi) www.mufi.info/fonts (sjekket 270907).

Mette Gismerøy Ekker
f. 1963
Kjeldeskriftavdelingen, Norsk Riksarkiv
mettegek@ulrik.uio.no

Christian-Emil Ore
førsteamanuensis, f. 1957
Institutt for lingvistiske og nordiske studier/EDD
Humanistisk fakultet
Universitet i Oslo, boks 1123 Blindern
NO-0317 Oslo
c.e.s.ore@edd.uio.no

