

NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Balancen i ordforrådet i en historisk ordbog	
Forfatter:	Ellert Þór Jóhannsson & Simonetta Battista	
Kilde:	Nordiska studier i lexicografi 16, 2023, s. 189–204	
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.

Balancen i ordforrådet i en historisk ordbog

Ellert Þór Jóhannsson & Simonetta Battista

This article is concerned with the process of selecting example citations to illustrate forms and meaning of the vocabulary registered in *A Dictionary of Old Norse Prose* (ONP), a historical dictionary project founded in 1939 and based in Copenhagen. After an initial phase of excerption of the relevant text material, the question arose as to whether the dictionary was lacking examples of some common words and expressions. To amend this potential deficit, it was decided early on to excerpt a few key texts extensively, by writing down every single word in context and adding it to the citation archive. In this article we present a case study that describes this endeavor and analyzes the results of this effort to even out a perceived imbalance in the lexical description. The results show that the additional citations unveiled some new definitions as well as phrases. However, by including hundreds of similar examples of common function words, such as conjugations, prepositions and pronouns, there arises a different kind of imbalance the user needs to be aware of.

KEYWORDS: historical dictionary, citations, excerption, dictionary editing

1. Indledning

I denne artikel fokuserer vi på materialet i en historisk ordbog, *Ordbog over det norrøne prosasprog* (ONP). ONP's leksikografiske arbejde med prosatekster fra middelalderen bygger på en samling af citater taget fra videnskabelige udgaver af bevarede håndskrifter efter filologiske principper. Formålet med artiklen er at analysere citatsamlingen i lyset af de leksikografiske udfordringer, man støder på under indsamlingen af materialet. Med udgangspunkt i ONP's selektive excerpering diskuterer vi det, som vi kalder balancen i beskrivelsen af ordforrådet, dvs. hvorvidt og hvor fyldestgørende de excerperede citater repræsenterer det bevarede tekstkorpus.

Efter en indledende fase med excerpering af det relevante tekstmateriale opstod spørgsmålet, om ordbogen manglede eksempler på en række almindelige ord og udtryk. Det førte til en særlig indsats, hvor fem udvalgte tekster blev excerperet ud fra et helhedsprincip, dvs. ord for ord, hvilket resulterede i ca. 38.000 ekstra citater. Vi vil her præsentere en case

study, hvor vi beskriver indsatsen og undersøger, hvordan den har påvirket balancen i det leksikografiske materiale, som i dag findes i ONP.

Artiklen er opbygget på følgende måde: efter denne indledning giver vi i andet afsnit et overblik over ordbogens principper og organiseringen af det leksikografiske materiale. I tredje afsnit beskriver vi den nævnte særlige excerperingsindsats. I fjerde afsnit analyserer vi materialet for at finde ud af, hvor mange og hvilke typer af ord, der blev excerperet fra de fem tekster. Vi sammenligner resultatet med tilgængelige lemmatiserede elektroniske tekster. Vi giver nogle eksempler på, hvordan det særligt excerperede materiale påvirker individuelle ordbogsartikler. Nogle konkluderende bemærkninger følger i femte afsnit.

2. *Ordbog over det norrøne prosasprog*: principper og organisering

Ordbog over det norrøne prosasprog er et ordbogsprojekt ved Københavns Universitet, som registrerer ordforrådet i oldnordiske prosatekster, overleveret i norske og islandske håndskrifter fra ca. 1150 til slutningen af middelalderen. ONP blev grundlagt i 1939 og var oprindeligt tænkt som en slags suppleringsværk til de store ordbøger fra det 19. århundrede, dvs. *Ordbog over det gamle norske sprog* af Fritzner (1886, 1891, 1896) og *An Icelandic-English Dictionary* af Cleasby & Vigfusson (1874). Man indså imidlertid hurtigt, at udarbejdelsen af en mere fyldestgørende beskrivelse af det norsk-islandske middelaldersprog ville kræve et nyt, selvstændigt opslagsværk. ONP blev baseret på en række filologiske principper, der adskilte den fra forgængerne: ordbogen skulle genspejle originalkilderne så nøjagtigt som muligt med unormaliseret ortografi og præcise kildehenvisninger; desuden skulle den omfatte eksempler på alle bevarede ord fra forskellige prosagenrer med ”særlig hensyntagen til ældste forekomst af såvel sproglige som syntaktiske fænomener og til kulturhistorisk interessant ordforråd” (Widding 1964:5). Et inspirerende projekt var *Middle-English Dictionary* som har en lignende filologisk tilgang til kildematerialet (cf. MED).

ONP har i tidens løb gennemgået forskellige stadier: et citatarkiv, en delvis trykt udgave (ONP 1–3), og siden 2010, et digitalt leksikografisk værktøj (ONP Online). Udgivelsen af trykte ordbogsbind blev indstillet i 2004, og kort derefter blev en stor del af det udarbejdede materiale gjort tilgængeligt på nettet (cf. Jóhannsson & Battista 2016). Onlineudgaven

er for nyligt blevet fuldstændigt redesignet og bliver løbende udvidet med nye artikler og links til andre relevante ressourcer og digitale hjælpemidler (cf. Jóhannsson, Battista & Wills 2021).

Det grundlæggende excerpingsarbejde var meget nøjagtigt og inkluderede alle værker fra alle bevarede prosagenrer. Resultatet blev et omfattende citatarkiv, der i dag består af over 800.000 repræsentative eksempler fordelt på ca. 65.000 lemmaer. Hele materialet er tilgængeligt og søgbart i ONP Online, dvs. både de færdige artikler med tilhørende citater, og de citater som ikke er redigeret endnu samt citater der af pladsmæssige årsager blev udeladt i de trykte bind. I modsætning til de trykte bind er der ingen pladsbegrænsning i onlineudgaven, hvilket har medført nye redigeringsretningslinjer: alle citater skal redigeres og placeres i strukturerede ordbogsartikler. På nuværende tidspunkt er over halvdelen af citatmaterialet bearbejdet i artikler.

ONP's kildemateriale er fordelt på 437 værker, som hver omfatter et varierende antal tekster. Et værk er defineret som enten samme tekst bevaret i mange forskellige håndskrifter, fx en islændingesaga som *Njáls saga*, der findes i mange håndskrifter i varierende versioner, eller urelaterede tekster, som har et fælles udgangspunkt, fx *Diplomatarium Islandicum*, der består af mange korte, juridiske dokumenter af forskellig slags (se Jóhannsson & Battista 2018 for nærmere diskussion om værker).

Værkerne fordeler sig videre på forskellige genrer. I alt er der ti prosagenrer defineret af ONP: islændingesagaer (fx *Njáls saga*, *Egils saga*); totter: korte fortællinger om islændinge; oldtidssagaer: heltesagn fra den germanske oldtid; samtidssagaer, der omhandler mennesker og begivenheder i forfatternes samtid (fx *Sturlunga saga*); høvisk litteratur: primært riddersagaer, dvs. tekster oversat fra oldfransk og originale værker inspireret af international litteratur; historisk litteratur, primært sagaer om de norske konger; lærde tekster, fx *Snorra Edda*, men også diverse tekster med videnskabeligt indhold; religiøs litteratur, omfattende fx tekster baseret på latinske, hagiografiske tekster og legendelitteratur oversat fra middelnedertysk og oldengelsk; juridiske tekster, dvs. norske og islandske lovttekster; diplomer: breve og officielle dokumenter, som belyser aspekter af dagliglivet i middelalderen.

I Tabel 1 findes en liste over alle genrer med angivelse af antallet af værker og citater under hver enkel (se også Jóhannsson & Battista 2018). Cirka 1 % af citaterne er ikke klassificeret under en genre, typisk fordi de

i streng forstand falder udenfor ONP's afgrænsede korpus, som fx personnavne eller poetiske ord.

TABEL 1. Fordeling af prosatekster i ONP's citatsamling.

Genrer	Citater	%	Værker	%
Religiøse tekster	213773	26,2 %	143	33 %
Historisk litteratur	139495	17,1 %	32	7 %
Islændingesagaer	97227	11,9 %	41	9 %
Høvisk litteratur	82592	10,1 %	52	12 %
Juridiske tekster	77032	9,4 %	50	11 %
Diplomer	61562	7,5 %	4	1 %
Samtidssagaer	58604	7,2 %	11	3 %
Oldtidssagaer	33971	4,2 %	29	7 %
Lærde tekster	24516	3,0 %	10	2 %
Totter	17659	2,2 %	65	15 %
Ukategoriseret	8867	1,1 %		
I alt	815.298	100 %	437	100 %

Man kan af Tabel 1 se, at de fleste citater stammer fra religiøse tekster, efterfulgt af historisk litteratur og islændingesagaer. Enkelte andre genrer omfatter også et betydeligt antal værker.

3. En særlig excerperingsindsats

Indsamlingen af materialet til ordbogsarbejdet stillede ONP's redaktører over for adskillige udfordringer. Excerpting af tekster var en selektiv proces, hvor redaktørerne skulle bedømme, hvilke citater, der var relevante ifølge ordbogens principper. Det betød samtidig, at interessen rettedes mere mod de mærkelige og usædvanlige former og vendinger, fremfor det almindelige. Efter første fase af excerptingen stod det endnu ikke klart for redaktørerne, hvorvidt det indsamlede materiale repræsenterede ordforrådet i kildeteksterne. Ved en nærmere gennemgang af citatsamlingen viste der sig at være en vis ubalance, idet man havde mange eksempler på mærkelige ord og former, men ikke særligt mange eksempler på ganske

almindelige ord, funktionsord, så som præpositioner, konjunktioner og pronominer, selvom disse ord forekommer hyppigt i tekstmaterialet. Ordbogens redaktion vurderede, at man formentlig gik glip af vigtige eksempler på sprogbrug og almindelige vendinger, som hørte til en udførlig leksikalsk beskrivelse af sproget. For at rette op på denne ubalance igangsatte man et særligt excerperingsprojekt, der skulle registrere hvert eneste ord fra nogle udvalgte, repræsentative, ældre tekster.

Dette blev muliggjort takket være eksterne ressourcer, som omtales i en tidlig rapport om ordbogsarbejdet.

På ganske uventet måde fik ordbogen øget arbejdskraft under krigen. Ved den såkaldte nødbeskæftigelse blev der af Socialministeriet stillet beløb til rådighed til at beskæftige arbejdsledige [...] De anviste medhjælpere kunne uden forkundskaber udskrive visse karakteristiske og centrale tekster efter thesaurusprincippet: hvert ord med sit citat på sin seddel [...] Alle disse sedler blev mærket på særlig måde (stemplet med et A i øverste venstre hjørne) [...] Særlig med hensyn til formord [dvs. funktionsord] eller småord har denne form for seddeludskrivning haft sin store betydning. (Widding 1964:7)

Fem tekster blev udvalgt ud fra følgende kriterier: der var tale om originale tekster, overleveret i forholdsvis gamle håndskrifter og tilgængelige i videnskabelige udgaver. Teksterne var i øvrigt særligt relevante fra et kulturhistorisk synspunkt. Det drejede sig om følgende tekster:

- *Gylfaginning* fra *Snorra Edda* (GKS 2367 4°, også kaldet *Konungsbók Snorra Eddu*, ca. 1300-1350): *Snorra Edda* er et enestående værk i nordisk middelalderlitteratur. Det er skrevet af Snorri Sturluson (1179-1241), en prominent forfatter og høvding i Island. *Gylfaginning* er en vigtig kilde til nordisk mytologi, idet størstedelen af vores viden om de nordiske guder er bevaret her.
- *Heiðarvíga saga* (Holm perg 18 4°, ca. 1300): en kendt islændingesaga, som er kun delvis bevaret i dette håndskrift. Håndskriftet blev tidligt delt op, og den del af det, som inkluderede første del af sagaen, gik tabt ved Københavns brand i 1728. Den anden del befandt sig i Sverige og overlevede derfor branden. Kun den del af sagaen, som er bevaret i den overlevede del af håndskriftet, er excerperet.

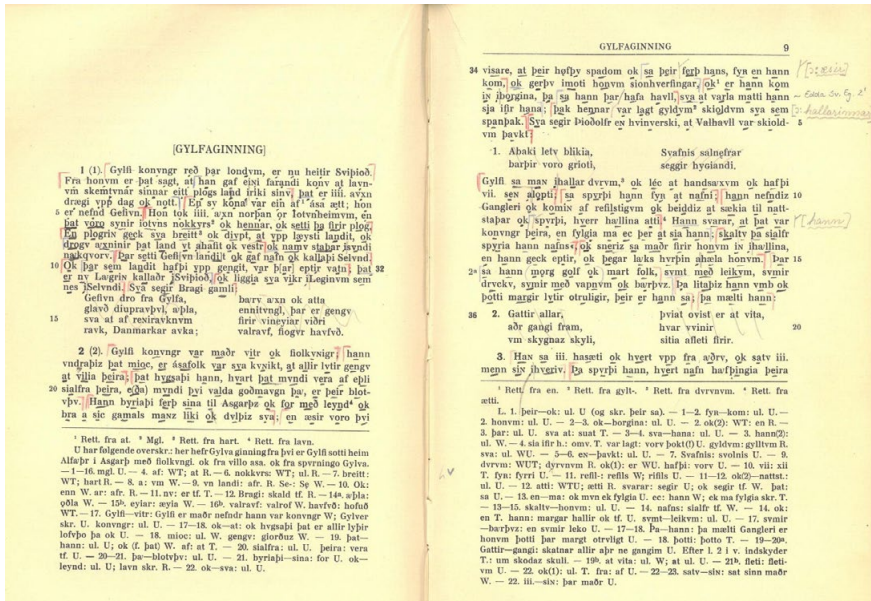
- et fragment af *Egils saga Skallagrímssonar* (AM 162 A 0 fol, ca. 1250): det ældste bevarede fragment af en af de mest berømte islændingesagaer.
- *Íslendingabók* (AM 113 folx, 1651): et enestående værk om Islands historie, forfattet af Ari Þorgilsson i 1100-tallet, men bevaret i et forholdsvis ungt papirhåndskrift, som imidlertid betragtes som en meget nøjagtig afskrift af en nu tabt skindbog, der formentlig skal dateres til ca. 1200.
- *Óláfs saga Tryggvasonar* (Holm perg 18 4°, ca. 1300): et vigtigt værk i islandsk litteraturhistorie, det første værk, der fokuserer på en enkelt konges historie. Sagaen findes i to versioner. Versionen i Holm perg 18 er forfattet af munken Oddr Snorrason i 1100-tallet. Denne tekst er bevaret i samme håndskrift som *Heiðarvíga saga*.

De fem tekster repræsenterer kun tre genrer, islændingesagaer (*Heiðarvíga saga* og *Egils saga*), historisk litteratur (*Íslendingabók* og *Óláfs saga*) og lærde tekster (*Gylfaginning*) (jf. Tabel 1). Det drejer sig om centrale genrer, der repræsenterer en betydelig del af det bevarede tekstkorpus. Samtidig bemærker man, at de udvalgte tekster ikke inkluderer oversættelser, juridisk sprog samt tekster fra genrer, som primært er forfattet i det 13. århundrede, fx høvisk litteratur og samtidsagaer.

I lyset af den senere udvikling i norrøn litteraturhistorie kan man måske argumentere for, at disse fem tekster blev fremhævet som de mest repræsentative for det norrøne ordforråd i en periode, hvor vægten blev lagt på bestemte genrer og karakteristikker. Teksterne repræsenterer forholdsvis velkendte og vigtige værker, der blev forfattet i Island allerede i 1100- og 1200-tallet. Måske har man betragtet dem som mere oprindelige eller tættere på højdepunktet i norrøn prosalitteratur end andre, og derfor som en slags guldstandard for sproget. En anden vigtig grund kan være, at udover at være relevante, var teksterne tilgængelige i pålidelige udgaver og forholdsvis nemme at transskribere for ”medhjælperen ... uden forkundskaber”, som det blev beskrevet i rapporten (Widding 1964:7).

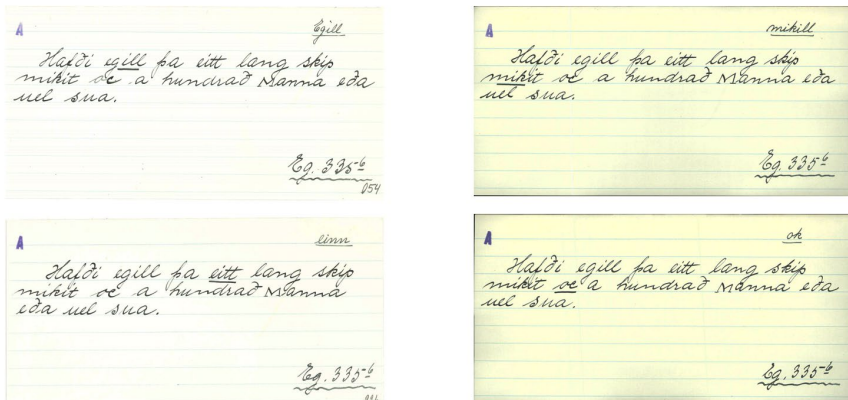
Excerperingen gik ud på at give eksempler på hvert eneste ord i sammenhæng. Tekststudgaverne blev gennemlæst af en redaktør, og hvert ord blev markeret som opslagsord (med understregning) i et citat (indesluttet

i hager) som skulle transskriberes på en seddel. Figur 1 viser et opslag fra den fuldt excerperede udgave af *Gylfaginning*.



FIGUR 1. Opslag fra en fuldstændig excerperet udgave, hvor alle ord er markeret som opslagsord (sort prik under ord) og citater er afgrænset (røde hager).

Resultatet af excerperingsarbejdet var en række sedler, ofte skrevet med samme hånd, hvor hvert citat forekommer under hvert eneste opslagsord, cf. Figur 2.



FIGUR 2. Fire sedler, skrevet med samme hånd, som viser samme citat under fire forskellige opslagsord.

Excerperingsindsatsen resulterede i tilføjelsen af 37.880 nye citater til ONP's citatsamling. Tallene kommer fra ONP's gamle arbejdsnoter, hvor man har registreret de forskellige faser af seddelopskrivningen. I næste afsnit diskuterer vi, hvilken indflydelse denne excerperingsindsats havde på balancen i citatmaterialet og hvordan den figurerer i det leksikografiske materiale, som i dag er tilgængeligt i onlineudgaven af ONP.

4. Undersøgelse og resultat

For at finde ud af hvordan excerperingsindsatsen har påvirket citatsamlingen har vi undersøgt de citater fra hver af de fem tekster som findes i dag i ONP's database. Resultatet vises i tabel 2. Det angives for hver tekst, hvor stor en andel de særligt excerperede citater udgør af det samlede antal citater i ONP.

TABEL 2. Særligt excerperede citater fra hver af de fem tekster sammenlignet med, hvor mange citater der findes i alt i ONP's database fra den pågældende tekst.

Tekst	Genre	Særlige citater	%	I alt i ONP	% af genre
Íslendingabók	Hist	3996	102 %	3899	2,8 %
Óláfs saga	Hist	4814	76 %	6279	4,5 %
Snorra Edda	Lær	15508	84 %	18483	75,0 %
Egils saga frg.	Isl	2986	29 %	10447	10,7 %
Heiðarvíga saga	Isl	10576	79 %	13439	13,8 %
I alt		37880		52547	

Tabel 2 viser, hvor mange citater, der blev excerperet fra hver af de fem udvalgte tekster i forhold til det samlede antal citater for hele den pågældende tekst. De fleste tekster var allerede blevet excerperet på normal vis. Man ville derfor forvente, at det samlede antal citater skulle være højere end antallet af citater fra de udvalgte tekster. To tal er dog lidt anderledes end forventet. Det lave tal for *Egils saga* skyldes, at kun et fragment af sagaen blev excerperet ord for ord. Tallene for *Íslendingabók* (102 %) tyder på, at den ikke har været excerperet i forvejen, og at der har været nogle dubletter, som så er blevet slettet under redigeringsprocessen. De 52.547 citater fra disse tekster svarer til 6,4 % af alle citaterne i ONP's database. De 16 hyppigste ord fra hver tekst vises i tabel 3.

TABEL 3. Tabellen viser de 16 hyppigste ord fra hver af de fem tekster med angivelse af ord, ordklasse og antal citater.

Snorra-Edda			Heiðarvíga saga			Egils saga			Íslendingabók			Ólafs saga		
Ord	Kl.	Cit.	Ord	Kl.	Cit.	Ord	Kl.	Cit.	Ord	Kl.	Cit.	Ord	Kl.	Cit.
ok	conj.	966	ok	conj.	674	ok	conj.	129	vera	vb.	172	ok	conj.	402
vera	vb.	663	sá	pron.	389	vera	vb.	101	ok	conj.	168	sá	pron.	208
hann	pron.	596	vera	vb.	383	sá	pron.	87	sá	pron.	140	hann	pron.	196
sá	pron.	511	hann	pron.	323	hann	pron.	75	er	conj.	120	vera	vb.	167
er	conj.	429	nú	adv.	187	hafa	vb.	61	en	conj.	109	er	conj.	104
þá	adv.	410	er	conj.	161	er	conj.	59	hann	pron.	105	í	præp.	81
en	conj.	311	til	præp.	160	en	conj.	57	þat	pron.	83	þá	adv.	73
í	præp.	234	hafa	vb.	131	til	præp.	54	þá	adv.	77	til	præp.	71
þat	pron.	223	þat	pron.	130	þú	pron.	51	í	præp.	73	koma	vb.	68
svá	adv.	217	barði	sb.	123	þat	pron.	50	inn	adv.	67	með	præp.	64
til	præp.	184	ek	pron.	122	þá	adv.	49	sonr	sb.	62	munu	vb.	59
segja	vb.	182	í	præp.	120	þar	adv.	40	hafa	vb.	60	ek	pron.	59
hafa	vb.	156	þú	pron.	117	svá	adv.	33	maðr	sb.	55	þat	pron.	58
heita	vb.	155	maðr	sb.	106	segja	vb.	33	á	præp.	53	þar	adv.	57
hon	pron.	145	koma	vb.	105	mál	sb.	33	til	præp.	48	maðr	sb.	53
allr	adj.	136	þar	adv.	101	í	præp.	31	vetr	sb.	37	sinn	pron.	51

De fleste ord er funktionsord af forskellige slags: primært konjunktioner, såsom *ok* 'og', *er* 'som' og *en* 'og, men', men også pronominer som *hann* 'han' og *sá* 'denne' og præpositioner. Der findes også andre ordklasser,

gængse verber som *vera* 'være', *segja* 'sige', *koma* 'komme', substantiver som *maðr* 'mand', og adverbier som *þá* 'da' og *nú* 'nu'.

En betydelig del af citaterne i ONP's materiale, der illustrerer funktionsordene og andre almindelige ord, er excerperet fra de fem udvalgte tekster. Tabel 4 viser de 16 hyppigste opslagsord fra de fem tekster, og hvor stor en procentdel citaterne derfra udgør af det samlede antal citater i ONP for det pågældende lemma.

TABEL 4. Opslagsord, antal citater og % af særlige citater.

Ord	Klasse	Fra de 5	ONP i alt	%
ok	<i>conj.</i>	2339	3230	72 %
vera	<i>vb.</i>	1486	3240	46 %
sá	<i>pron.</i>	1335	2298	58 %
hann	<i>pron.</i>	1295	1694	76 %
er	<i>conj.</i>	873	3087	28 %
þá	<i>adv.</i>	703	1291	54 %
en	<i>conj.</i>	579	796	73 %
þat	<i>pron.</i>	544	1053	52 %
í	<i>præp.</i>	539	1288	42 %
til	<i>præp.</i>	517	1669	31 %
hafa	<i>vb.</i>	454	2475	18 %
ek	<i>pron.</i>	368	745	49 %
maðr	<i>sb.</i>	344	1325	26 %
koma	<i>vb.</i>	326	3340	10 %
þú	<i>pron.</i>	308	444	69 %
nú	<i>adv.</i>	294	684	42 %

Mellem halvdelen og tre fjerdedele af eksemplerne, som indeholder hyppigt brugte konjunktioner og pronominer, kommer fra disse tekster, fx *ok* = 72 %, *hann* = 76 %, *en* = 73 %. Samtidig kan man se, at den selektive excerpering af mere betydningsfulde ord alligevel har været "fornuftig". Fx har *maðr* proportionelt flere citater fra andre værker end de fem udvalgte. Det samme kan man sige om verber som *hafa*, *koma* og *vera*, som har mange betydningsnuancer.

For at sætte resultatet fra undersøgelsen ind i videre sammenhæng kan man sammenligne det med moderne ressourcer. Der findes en del elek-

troniske tekstudgaver, hvor enkelte norrøne tekster er blevet fuldstændigt transskriberet fra et enkelt håndskrift og lemmatiseret. Mange af disse tekster er tilgængelige online i *Arkiv for nordiske middelaldertekster* (MENOTA), der samler elektroniske udgaver af norrøne tekster og giver adgang til dem og relateret data. Resultatet af vores undersøgelse kan sammenlignes med et antal eksempler fra relevante tekster i MENOTAs arkiv. Tabel 5 viser en oversigt over de mest almindelige ord i de norrøne tekster som findes i MENOTA.

TABEL 5. De 16 hyppigst forekommende ord i elektroniske udgaver af norrøne tekster (MENOTA).

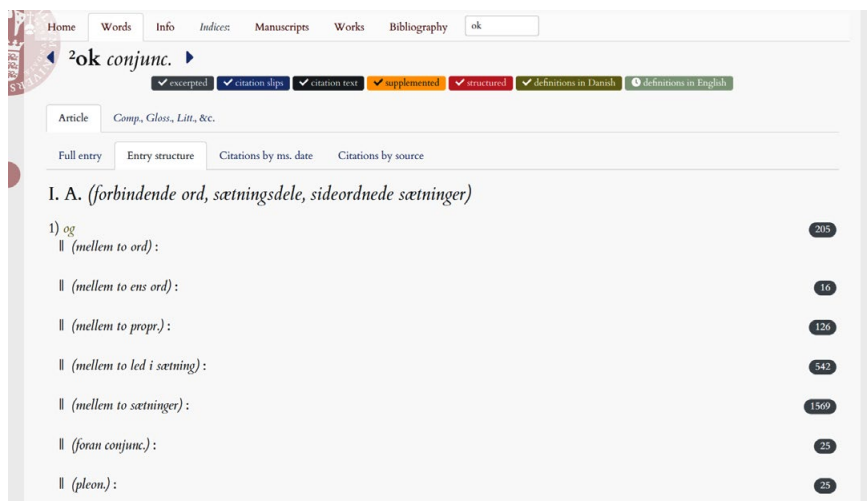
Ord	Ordklasse	Antal
ok	<i>conj.</i>	26154
hann	<i>pron.</i>	21810
sá	<i>pron.</i>	21757
vera	<i>vb.</i>	13672
er	<i>conj.</i>	11135
at	<i>conj.</i>	8930
í	<i>prep.</i>	7517
þá	<i>adv.</i>	6729
til	<i>prep.</i>	6485
hafa	<i>vb.</i>	5809
með	<i>prep.</i>	5573
sem	<i>conj.</i>	5066
ek	<i>pron.</i>	5028
sinn	<i>pron.</i>	4858
svá	<i>adv.</i>	4810
allr	<i>adj.</i>	4803

Når man sammenligner tabel 5 med tabel 4, er det stort set de samme ord som figurerer begge steder, fx konjunktionen *ok*, verbet *vera* og pronomenet *hann*. Den største forskel er, at man i ONP har pronomenet *þat* 'det', som ikke forekommer som selvstændigt lemma i MENOTA-teksterne.

Grunden hertil er, at mange tekstudgivere har valgt at lemmatisere *þat* som en bøjningsform under lemmaet *sá*, som *þat* i mange tilfælde er neutrumsform af.

Man lægger også mærke til, at der i MENOTA-teksterne forekommer nogle ord med mange eksempler, som ikke er lige så godt dokumenteret i ONP's data. Det er især tilfældet ved konjunktionen *at*, for hvilken man ikke finder så mange excerperede citater, selvom man ville have forventet, at ordet optrådte hyppigt i de fem udvalgte tekster. Når man ser nærmere på dataene, viser det sig, at konjunktionen *at* er blevet skrevet op på sedler på samme måde som alle andre ord fra de fem tekster. Uoverensstemmelsen skyldes at *at* blev redigeret i det første bind af den trykte udgave (ONP 1). Den gang var praksis ved redigeringsarbejdet, at alle citatsedlerne blev indtastet i ordbogens database, men kun et udvalg blev brugt ved den endelige redaktion af opslagsord med mange eksempler. Ved konjunktionen *at* har man afvejet fra denne fremgangsmåde, idet man tidligt i redigeringsprocessen indså, at der var disproportionalt mange eksempler på ordet. Derfor blev ikke alle citater tastet ind, men derimod kun et udvalg. Samme fremgangsmåde blev brugt for andre funktionsord tidligt i alfabetet, fx præpositionerne *af* 'af' og *á* 'på'.

Alle citater, som hører under opslagsord, der er blevet redigeret efter ONP 1–3 udkom, er til gengæld blevet en del af ONP's database, og de er blevet redigeret eller skal redigeres i de færdige artikler. Til sammenligning kan man se, at konjunktionen *ok* ikke er blevet reduceret på samme måde som *at*, idet den har 3230 eksempler, som alle sammen indgår i en struktureret ordbogsartikel. Figur 3 illustrerer delvis artiklens struktur. Det er angivet, hvor mange citater, der hører til hver betydning.



Home Words Info Indices Manuscripts Works Bibliography ok

◀ **ok conjunc.** ▶

✓ excerpted ✓ citation slips ✓ citation text ✓ supplemented ✓ structured ✓ definitions in Danish 0 definitions in English

Article Comp., Gloss., Lit., &c.

Full entry Entry structure Citations by ms. date Citations by source

I. A. ((forbindende ord, sætningsdele, sideordnede sætninger))

1) og 205

|| (mellem to ord): 16

|| (mellem to ens ord): 126

|| (mellem to propr.): 542

|| (mellem to led i sætning): 1569

|| (mellem to sætninger): 25

|| (foran conjunc.): 25

|| (pleon.): 25

FIGUR 3. Første del af artiklen *ok* i ONP, hvor der kun vises struktur og antal citater.

På Figur 3 kan man se, at der findes 1569 citater, som viser brugen af *ok* ”mellem to led i sætning”. Man må konstatere, at den fuldstændige excerpering af tekster og inklusion af alle citater i ordbogens database også kan give et skævt billede af ordforrådet med en overvægt af ensartede eksempler.

Der er dog nogle tilfælde, hvor den særlige excerperingsindsats resulterede i nye betydninger og fraser. Figur 4 viser eksempler fra artiklen *mikill* adj. ’megen, stor’, hvor den særlige excerpering afslørede både en ny betydning og nogle fraser, som ikke var registreret tidligere. Fra *Óláfs saga* har man et eksempel, hvor adjektivet refererer til stærkt lys og fra *Heiðarvíga saga* nogle vendinger som *ríða mikinn dyn* ’ride med meget støj’ og *gera mikit af sér* ’gøre en stor indsats på egen hånd’, der illustrerer dette.



7) [til e-ts/e-s / um e-n] hárd, kraftig, voldsom, barsk

• **ríða mikinn dyn**

ok ríða þeir heim **mikinn** dyn itunit eptir hæðom uelli Heið 71¹

• **gera mikit af sér**

ok nv beriaz þeir allir. ok gera **mikit** af sér. ok þar falla þeir synir Eids Heið 93⁴

19) (om lys) klar, stærk

Hit se ek at aþessi ... tíþ er borinn i Norege konungs s. (: sarr) með biortvnn fylgiom ok er **mikit** lios yfir honum ÓTODS 20²

FIGUR 4. Udvalgte dele af artiklen *mikill* adj. ’megen, stor’ fra ONP.

Disse eksempler viser, at den særlige excerperingsindsats kunne resultere i mere nøjagtig leksikalsk beskrivelse. Dette er dog ikke tilfældet ved de fleste funktionsord.

5. Konklusion

I denne artikel har vi fokuseret på ONP's citatsamling i lyset af de leksikografiske udfordringer, man støder på under indsamlingen af materiale og diskuteret ordforrådet fra de fem særligt udvalgte tekster og hvordan de figurerer proportionelt i materialet og i individuelle ordbogsartikler. Formålet med den særlige excerperingsindsats var at rette op på en formodet ubalance i citatsamlingen med henblik på en mere fyldestgørende leksikalsk beskrivelse. Denne case study har vist at den særlige excerperingsindsats indbragte flere eksempler på underrepræsenterede ord, fx konjunktioner, pronominer, og gængse verber. I flere tilfælde resulterede det dog i mange unødvendige eksempler på samme slags sprogbrug og samme betydning af helt almindelige ord. Man har også kunnet dokumentere nye betydninger og vendinger, omend ikke så mange, som man måske havde håbet på, da man gik i gang med at supplere citatsamlingen på denne måde. Med forsøget på at rette op på den antagne ubalance opstod der en anden slags skævhed i form af et overvældende antal ensartede citater, som ordbogsbrugerne skal være opmærksomme på.

ONP's forsøg med fuld excerpering af de fem tekster kunne fra et moderne synspunkt betragtes som en slags korpuslingvistik, dvs. at inkorporere alle ord fra udvalgte tekster i ordbogen. Det er imidlertid uhensigtsmæssigt at anvende denne metode, hvis man fokuserer på den leksikalske beskrivelse af sproget, fordi datamængden bliver uoverskuelig (som vi har set i forbindelse med artiklen *ok*). Materialet skal afgrænses med henblik på overskuelige, brugervenlige ordbogsartikler, som samtidig skal dække alle nuancerne i sproget. ONP Online viser alle citater, som findes i citatsamlingen, men dette princip skal muligvis revideres i forbindelse med beskrivelsen af funktionsord og andre almindelige ord som kan have hundredvis af eksempler.

Inkorporeringen af korpusmateriale er dog muligt med sprogteknologiske værktøjer og kan med fordel bruges til at supplere ordbogsdata. Man kan benytte elektroniske udgaver af tekster og andre digitale ressourcer til at give brugerne mere fyldestgørende oplysninger om ordforrådet også

til andre formål end at slå en betydning op, fx overblik over mulige grammatiske former, forekomst i andre ordbøger, forskellige tekstudgaver osv. Tilknytning til eksterne ressourcer er noget som ONP arbejder på sideløbende med redigeringsarbejdet, og det er blevet diskuteret andre steder (fx Wills, Jóhannsson & Battista 2018, 2021 og Wills & Jóhannsson 2019). Supplerende citater kan godt stilles til rådighed for brugerne, men et højt antal eksempler kan virke forstyrrende på artikelstrukturen.

Litteratur

- Cleasby, Richard & Guðbrandur Vigfússon. 1874. *An Icelandic-English Dictionary*. Oxford: Clarendon Press.
- Fritzner, Johan 1886, 1891, 1896. *Ordbog over Det gamle norske Sprog 1-3*, rev. udg. Kristiania: Den norske forlagsforening.
- Jóhannsson, Ellert Þór & Simonetta Battista 2018. Middelaldertekster som sproglig ressource. I: Ásta Svavarsdóttir & Helga Hilmisdóttir (red.) *Nordiske Studier i Leksikografi 14: Rapport fra 14. Konference om Leksikografi i Norden, Reykjavík 30. maj-2. juni 2017*, 152-161.
- Jóhannsson, Ellert Þór & Simonetta Battista 2016. Ordbog over det norrøne prosasprog Online: struktur og brug. I: Gudiksen, Asgerd & Henrik Hovmark (red.) *Nordiske Studier i Leksikografi 13: Rapport fra 13. Konference om Leksikografi i Norden, København 19.-22. maj 2015*, 165-175.
- Jóhannsson, Ellert Þór, Simonetta Battista og Tarrin Wills 2021. Legacy data in a digital age. I: Reinsone, Sanita, Inguna Skadiņa, Jānis Dau-gavietis & Anda Baklāne (red.) *Digital Humanities in the Nordic Countries 2020, Post-Proceedings of the 5th Conference Digital Humanities in the Nordic Countries Riga, Latvia, October 21-23, 2020*, 248-254.
- MED = *Middle English Dictionary*. Ed. Robert E. Lewis, et al. Ann Arbor: University of Michigan Press, 1952-2001. Online edition in *Middle English Compendium*. Ed. Frances McSparran, et al. Ann Arbor: University of Michigan Library, 2000-2018. <<http://quod.lib.umich.edu/m/middle-english-dictionary/>>. Hentet marts 2023.
- MENOTA = *Medieval Nordic Text Archive*. *Arkiv for nordiske middelaldertekster*. <<https://menota.org>>. Hentet august 2022.
- ONP 1-3 = Degnbol, Helle, Bent C. Jacobsen, James E. Knirk, Eva

- Rode, Christopher Sanders & Þorbjörg Helgadóttir (red.) *Ordbog over det norrøne prosasprog / A Dictionary of Old Norse Prose. ONP Registre* 1989. ONP 1: *a–bam* .1994. ONP 2: *ban–da*. 2000. ONP 3: *de–em*. 2004. ONP *Nøgle/Keys*. 2004. København: Den Arnamagnæanske Kommission.
- ONP Online = *Ordbog over det norrøne prosasprog / A Dictionary of Old Norse Prose*. <<http://onp.ku.dk>>. Hentet august 2022.
- Widding, Ole 1964. *Den Arnamagnæanske Kommissions Ordbog, 1939–1964: Rapport og plan*, København: G.E.C.GADS Forlag.
- Wills, Tarrin & Ellert Þór Jóhannsson 2019. Reengineering an Online Historical Dictionary for Readers of Specific Texts. I: Kosem, Istok, (red.) *Electronic lexicography in the 21st century: Smart lexicography: Proceedings of the eLex 2019 conference*. Brno: Lexical Computing CZ, 116–129.
- Wills, Tarrin, Ellert Þór Jóhannsson & Simonetta Battista 2018. Linking Corpus Data to an Excerptbased Historical Dictionary. I: Čibej, Jaka, Vojko Gorjanc, Istok Kosem & Simon Krek (red.) *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*. Ljubljana: Ljubljana University Press, Faculty of Arts, 979–987.
- Wills, Tarrin, Ellert Þór Jóhannsson & Simonetta Battista 2021. Integrating TEI/XML Text with Semantic Lexicographic Data. I: Reinsone, Sanita, Inguna Skadiņa, Jānis Daugavietis & Anda Baklāne (red.) *Digital Humanities in the Nordic Countries 2020, Post-Proceedings of the 5th Conference Digital Humanities in the Nordic Countries Riga, Latvia, October 21–23, 2020*, 16–25.