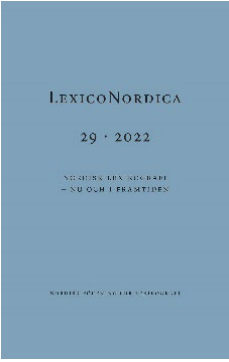


# LexicoNordica

Titel:	En historisk ordbogs digitale fremtid	
Forfatter:	Tarrin Wills	
Kilde:	LexicoNordica 29, 2022, s. 175-194	
URL:	<a href="http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive">http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive</a>	

© 2022 LexicoNordica och författarna

## 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.

# En historisk ordbogs digitale fremtid

*Tarrin Wills*

The Dictionary of Old Norse Prose (ONP) is an ongoing historical project covering an important medieval language and its literature. The dictionary is based on traditional methods but has amassed a wealth of resources as part of its long work. In order to stay relevant, ONP has been developed with a view to improve its editing methods, expand its user base and connect to other digital resources that have become available in more recent times. This paper presents a number of methods by which we achieve these ends.

## 1. Indledning

Det tager lang tid og kræver mange ressourcer at udarbejde store historiske dokumentationsordbøger. Man hører ofte om sådanne udfordringer, som for eksempel i denne opsummering af historien om *Middle English Dictionary* (MED) på projektets hjemmeside (<quod.lib.umich.edu/m/middle-english-dictionary/about>):

The 70-year Middle English Dictionary project drew on support from (and probably exhausted the patience of) many agencies and individuals over that span, both internal and external, to an extent that can only be hinted at here.

Mange lignende tilfælde kendes fra andre ordbøger, eksempelvis *Oxford English Dictionary* (se f.eks. Mugglestone 2005:33-34). Da projekterne løber over en lang periode, er man nødt til løbende at opdatere de redaktionelle metoder når faget udvikler sig, og når der kommer nye standarder og flere eksterne ressourcer i form af nye opslagsværker, tekstudgaver o.l.

Denne artikel beskriver nogle tiltag der sigter mod at løse de udfordringer der er forbundet med langsigtede historiske ordbogsprojekter, især hvad angår tilgængelighed af ikke-redigeret materiale og redaktionsprocessens hastighed. De valgte løsninger gør brug af nye teknologier og eksterne ressourcer, og i mange tilfælde er resultaterne målbare.

## 2. Baggrund

*Ordbog over det norrøne prosasprog* (ONP) er et ordbogsprojekt ved Københavns Universitet, der blev grundlagt i 1939. ONP registrerer ordforrådet i oldnordiske prosatekster overleveret i norske og islandske håndskrifter fra ca. 1150 til slutningen af middelalderen. ONP bruger MED som model for sin redaktion, da MED på ordbogens begyndelsestidspunkt var den metodisk nyeste og bedste store historiske ordbog over et lignende gammelt sprog. I 80'erne begyndte projektet at bruge databaser til at samle og organisere dataene, og de trykte bind, der udkom 1989-2004, blev eksporteret direkte fra ONP's database.

Ifølge ONP's redaktionelle principper skal materialet afspejle originalkilden så nøjagtigt som muligt, hvilket betyder at citaterne er taget fra videnskabelige udgaver eller direkte fra håndskrifter. I de fleste tilfælde henviser ONP til både tekstudgaver og de håndskrifter som udgaverne bygger på (se f.eks. Battista og Jóhannsson 2014, Jóhannsson og Battista 2016).

I 1989 udkom ordbogens registerbind der fungerer som et selvstændigt opslagsværk over håndskriftskilder, videnskabelige tekstudgaver og relevant litteratur. Registerbindet blev efterfulgt af tre trykte bind med ordbogsartikler, som dækker alfabetet fra a- til em-. Efter 2004 blev udgivelsen af det trykte værk indstillet, og det blev besluttet at gøre ordbogsmaterialet tilgængeligt på nettet, i første omgang ved at scanne det materiale som endnu ikke var

udgivet. I 2010 blev onlineudgaven lanceret med både materiale fra de udgivne bind og andet materiale som primært bestod af lemmalisten samt tilhørende citater i form af scannede sedler fra ONP's arkivsamling. En ny onlineudgave blev lanceret i 2019, og denne giver adgang til flere materialer i ONP's database og forbinder ONP til en række eksterne ressourcer.

Det norrøne sprog og dets litteratur er vigtige for mange forskellige forskningsområder og fag. De norrøne tekster indeholder i de fleste tilfælde de tidligste belæg på en stor del af ordforrådet i de skandinaviske sprog, og teksterne er derfor meget centrale i sproghistorisk sammenhæng. Litteraturen omfatter mange betydningsfulde genrer og værker, f.eks. sagaerne og eddadigtningen. Kongesagaerne og andre tekster bruges som centrale kilder til Nordens historie i middelalderen, og de mange lovtekster er særligt interessante for lovhistorie og undersøgelser af sociale normer og værdier i middelalderen. Også Snorres Edda og den poetiske Edda er centrale kilder i vores forståelse af den førkristne religion og mytologi i Norden og andre steder. Der findes i dag mange gode oversættelser af de vigtigste værker, men historikere, religionshistorikere og andre forskere der ikke selv er sprogforskere, er stadig nødt til at kunne læse de oprindelige kilder for at kunne kontrollere deres tolkninger og for at kunne henviser til de oprindelige kilder.

Videreudviklingen af ONP har flere formål. Vi vil gøre de mange digitale ressourcer der allerede er blevet samlet, mere tilgængelige samt understøtte de forskellige forskningsprojekter der har brug for at tilgå og udnytte ordbogens oplysninger. Ordbogen kan også bruges som hjælp til læsning af norrøne tekster der ikke er blevet oversat. Vi vil gøre ordbogen mere kompatibel med nye leksikografiske metoder og med andre ressourcer, især korpuslingvistiske og leksikografiske ressourcer og digitale udgaver af norrøne tekster. Målet er at gøre redaktionsarbejdet mere effektivt, så ressourcerne kan bruges på at forbedre ordbogens indhold og gøre ordbogen lettere tilgængelig.



### 3. Forbedring af redaktionsprocessen

Allerede fra 1980'erne blev ONP redigeret i et databasesystem (DBase og senere Oracle), og den daværende redaktør Bent Chr. Jacobsen udviklede ONP's redigeringsprogram over mange år. Teknologien i systemet er avanceret nok til at forbinde ordbogens citater, artikler og registrene over udgaver og håndskrifter med hinanden, og alle dele kan redigeres gennem redigeringsprogrammet der kører i Windows-operativsystemet. Det tager dog mange måneder inden en ny redaktør kender systemet godt nok til at kunne skrive en ordbogsartikel, og det kræver mange års erfaring inden man behersker det fuldstændigt.

#### 3.1. Redigering af ordbogen

Det gamle redigeringsystems form og arbejdsgange stammer fra 1980'erne hvor man var nødt til at redigere direkte i databasens tabeller. Lemmaer, definitioner og citater m.m. kan eksporteres til et kontrolformat, bogformat eller web. Så længe ordbogen har flere meget erfarne redaktører der arbejder effektivt i det gamle system, giver det ikke mening at lave det om. Men nye teknologier gør det nemmere at udvikle mere brugervenlige grænseflader, og derfor har ordbogen udviklet et nyt redigeringsystem der kører parallelt med det gamle som stadig er i brug. Det nye system bevarer de gamle strukturer, og det bruger det samme databasesystem som det gamle system.

Det nye system er integreret i ONP's website. Det giver mulighed for at redigere ordbogen direkte på websitet: Man kan tilføje definitioner, flytte citater rundt mellem forskellige afsnit og definitioner og se resultaterne med det samme og i det samme format som brugeren ser dem i. Det betyder at nye redaktører meget hurtigere kan lære at arbejde med redigeringsystemet. De to systemer kører parallelt, men der er stadig problemer med at implementere

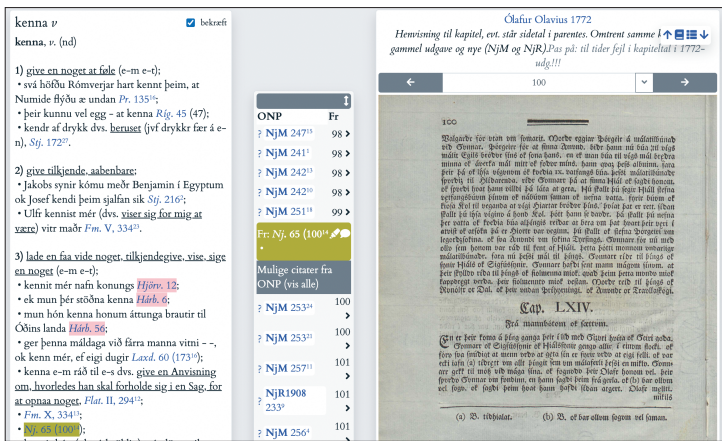
den nuværende datastruktur i to forskellige applikationer. Til gengæld er det meget nemmere i en webapplikation at oprette links til andre nyttige ressourcer for redaktionen, f.eks. til ordbøger og korpusser på nettet. I det lange løb skal strukturen laves om så den bedre afspejler de nyere digitale standarder for ordbøger (f.eks. Text Encoding Initiative (TEI) og OntoLex (Bosque-Gil & Gracia 2019)). De nye standarder bruger eksplicitte relationer mellem artiklens dele i selve data, i modsætning til ONP's datastruktur hvor de digitale relationer ofte er implicite.

### 3.2. Supplering fra ældre ordbøger

En af de processer som har haft størst gavn af webapplikationens nye muligheder, er suppleringen med materiale hentet fra ældre ordbøger. ONP supplerer sine citater fra de mest centrale ældre ordbøger, især *Fritzner* og *Clv*. Processen er tidskrævende: Man tager hvert citat i den gamle ordbog og slår op i den tekstudgave som der henvises til i citatet. Den der supplerer, prøver så at finde det tilsvarende tekststykke i den udgave som ONP bruger. Hvis det lykkes, og citatet ikke allerede findes i databasen, tilføjer man det til databasen med de korrekte henvisninger til udgaven og ordet så det supplerede citat kan bruges i forbindelse med redigeringen af ordet. Gennemsnitligt er ca. 9 % af citaterne til et opslagsord i ONP kommet ind gennem denne suppleringsproces. Selvom ordbogen har adgang til flere digitale og fysiske ressourcer, vurderer vi dog at man stadig bruger ca. 20 % af den samlede redigeringsstid på denne gamle måde at supplere på.

En ny digital redigeringsproces i webprogrammet gør arbejdet langt nemmere og hurtigere. Digitale udgaver af de vigtigste gamle ordbøger bliver behandlet automatisk og henvisningerne gøres klikbare. Henvisninger der ikke er relevante for ONP, f.eks. fordi de henviser til poetiske eller for unge tekster (jf. ordbogens register), fremhæves. Når man klikker på en henvisning, henter pro-

grammet citererne fra artiklen der hører til det samme værk, samt henvisninger til citater der ligger i nærheden af den gamle ordbogs citat i den gamle udgave. Den gamle udgaves scannede side vises ved siden af så man hurtigt kan se citatets kontekst. Såfremt citatet allerede findes i artiklen, kan man markere det med et klik i et felt og dermed registrere overensstemmelsen. Findes citatet ikke i databasen, kan man klikke på et citat fra et andet ord i nærheden. Det åbner en side i den nye udgave, hvor alle citater på siden vises (se afsnit 5.1). Redaktøren kan blade mellem siderne, finde citatet i den nye udgave og tilføje citattekst samt linjehenvisning. I tilfældet nedenfor (figur 1), fra *Fritzner*, forbinder systemet de nye oplysninger med den gamle ordbogs digitale tekst.



Figur 1: Suppleringsystemet med den gamle ordbogsartikel, citater fra ONP og billede af den udgave som den gamle ordbog henviser til.

Den nye suppleringsfunktion forbinder flere digitale ressourcer:

- gode digitale versioner af de gamle ordbøger
- et gammelt, men nyttigt digitalt register over de gamle ordbøgers litteratur
- billeder af de gamle udgaver, som nu alle er uden ophavsret, og som i de fleste tilfælde findes på nettet

- tusindvis af henvisninger fra supplerede citater i ONP til de gamle ordbøgers henvisninger, som tidligere blev indtastet manuelt
- visning i websitet, der samler alle citater der er blevet excerperet fra en side i originalværkets udgave, sammen med et billede af siden, og gør citaterne redigerbare for redaktorerne (jf. figur 3, afsnit 5.1).

Dette system gør processen mindst dobbelt så hurtig som den gamle, og det er simpelt nok til at en studentermedhjælper kan lære at bruge det. Det samler samtidig flere oplysninger om forbindelsen mellem ONP og de ældre ordbøger og gør det nemmere at finde citaterne herefter.

#### 4. Tilgængelighed og formidling

I de trykte ordbogsbind (a- til em-, ca. 20 % af det samlede antal citater) bruges både engelsk og dansk som målsprog, men siden 2010 er langt de fleste definitioner kun blevet skrevet på dansk. I skrivende stund udgør det allerede redigerede materiale 55 % af alle citater, og heraf mangler en stor del altså engelske definitioner. Derfor er der store huller i ordbogens semantiske dækning af ordforrådet, især hvad angår den semantiske analyse med engelsk som målsprog. Trods de manglende engelske definitioner vil ONP sikre at ordbogen er så brugbar som mulig så man kan anvende den som primært udgangspunkt og ikke behøver at konsultere ældre ordbøger.

ONP bruger to metoder til at tilføje de manglende semantiske analyser midlertidigt: supplerung fra ældre ordbøger på engelsk og dansk/norsk, og automatisk oversættelse udført af Google Oversæt.

#### 4.1. Supplering af uredigerede ord

ONP's digitale ordliste indeholder detaljerede henvisninger til 36 andre historiske ordbøger, glossarer og ordlister over det norrøne sprog. I mange tilfælde er disse ordbøger blevet digitaliseret, især de vigtigste (*Fritzner* og *Clv*; se ovenfor i afsnit 3.2), hvilket åbner muligheden for at sammenkæde dem digitalt med ONP's artikler.

For at linke et ord i ONP med de ældre ordbøger bruger webprogrammet henvisninger fra databasen til at rekonstruere det tilsvarende ord i de ældre ordbøger ved at konvertere ortografien. Henvisningerne til de ældre ordbøger i ONP er imidlertid udviklet til den trykte udgave af ordbogen. De indeholder ofte kun dele af opslagsordet i den ordbog der bliver henvist til, for at vise de ortografiske forskelle mellem ONP's opslagsord og opslagsordet i den ældre ordbog. Det er derfor vanskeligt at rekonstruere opslagsordet alene ud fra disse data således at man kan søge efter ordet i en digital version af den ældre ordbog. Programmet forsøger at rekonstruere opslagsordet alligevel, hvilket lykkes i langt de fleste tilfælde, og derfor kan webprogrammet vise de relevante ordbogsartikler fra de andre ordbøger på den samme webside som ONP's artikel. I de tilfælde hvor det ikke lykkes, kan redaktørerne hente de rigtige artikler i andre digitale ordbøger og manuelt linke dem til ONP's artikel. Brugere kan derfor få adgang til indholdet i de ordbøger som der ikke kan linkes til automatisk.

Hvad angår artikler/ord i *Fritzner*, *Clv*, *Hertz* og ordbøgerne fra *málið.is* (en fælles portal til flere islandske ordbøger), kan ONP vise digitale artikler af høj kvalitet fra disse ordbøger. Det samme system der bruges af ONP til at supplere fra disse ordbøger (se afsnit 2.2), bruges også til at lade en ekstern bruger læse de tekster som de gamle ordbøger henviser til. Ved *Lexicon Poeticum* (*LP*) viser websiden et scannet billede af ordbogens side.

For dem der helst vil læse definitioner på engelsk, har *Clv* engelsk som målsprog. Herudover trækker webprogrammet på en

række andre ressourcer til at supplere det engelske. Det henter via en Application Programming Interface (API) oplysninger fra The Skaldic Project (Skaldeprojektet) som er et udgivelsesprojekt der lemmatiserer alle ord med henvisning til ONP. En API gør det muligt at forbinde et eksternt program (en webapplikation for eksempel) til et andet program. I dette tilfælde kan man søge på et lemma og derefter hente en detaljeret konkordans af ord med det samme lemma i korpusset. I The Skaldic Project oversætter man teksterne til engelsk ord for ord så vidt det er muligt, og der findes nu oversættelser af 150.000 ord med 11.000 lemmaer. Det betyder at ONP for mange artiklers vedkommende kan hente alle citater fra The Skaldic Project's hjemmeside sammen med alle oversættelser af det enkelte lemma. Disse oversættelser supplerer ONP's danske eller manglende engelske definitioner.

Med disse eksterne ressourcer kan ONP supplere sit materiale med de manglende oplysninger om ordets betydninger og kontekst uden at brugeren er nødt til at forlade sitet. Det supplerende materiale giver oplysninger på både engelsk, dansk, norsk og islandsk. For mange brugere er engelsk dog det eneste relevante målsprog. I disse tilfælde bruger vi automatisk oversættelse.

## 4.2. Engelsk

Ifølge Google Analytics har kun 13 % af ONP's brugere svensk, dansk eller norsk som standardsprog i deres browser. 20 % af brugerne besøger ONP fra Danmark, Norge og Sverige, og 12 % fra Island hvor de fleste sandsynligvis kan læse dansk. Det betyder at højst en tredjedel af ordbogens brugere kan antages at kunne læse dansk (se Wills, Battista & Jóhannsson 2021) selvom 60 % af det semantisk redigerede materiale kun har dansk som målsprog, ligesom flere af de eksterne ressourcer.

Som en midlertidig løsning bruger ONP Google Oversæt til at oversætte danske og norske definitioner flere steder. Oversættelsen

sker når brugeren klikker på et stykke norsk eller dansk tekst som er blevet fremhævet. Det betyder at søgemaskiner og webarkiver ikke ser den automatisk oversatte tekst, hvilket ellers kunne antyde at det er ONP's oversættelse. Oversættelsen forsynes med tydelig kildeangivelse, og brugeren kan se den oprindelige tekst (jf figur 2). I en undersøgelse (Wills 2021) vurderede vi at omkring 90 % af de brugeranmodede oversættelser var tilstrækkelige til at kunne hjælpe brugeren.

Denne funktion bruges især til ONP's definitioner på dansk, men den er ydermere blevet anvendt på de linkede eksterne ressourcer der har digital tekst på dansk og norsk, dvs. *Fritzner*, *Blöndal* og *Hertz*. Når man henter disse artikler fra eksterne ressourcer, gør webapplikationen teksten klikbar så brugeren kan anmode om en oversættelse fra Google Oversæt.

Figur 2: Brugeranmodet automatisk oversættelse af en dansk definition.

## 5. Forbindelse med andre ressourcer

Når ordbogen kan forbindes med andre ressourcer, er det også muligt at bruge den til andre formål. ONP er ikke kun en lingvistisk ressource: Mange bruger ordbogen til at kunne læse tekster der ikke er blevet oversat. Andre bruger den til at undersøge sociale og kulturelle fænomener ved at slå ord op der bruges i forbin-

delse med sådanne fænomener så de kan finde forskellige citater og kilder til disse fænomener.

ONP har sammenkædet mange typer af data i løbet af de sidste årtier som en del af redaktionsarbejdet. Nogle af disse henvisninger refererer til traditionelle ressourcer: tidligere ordbøger, håndskrifter, udgaver og paralleller (kilder). Andre henviser til nye, bl.a. begrebsordbøger (tesaurusser) og andre ressourcer der indeholder semantiske oplysninger om ordene og korpusset.

### 5.1. Tovejshenvisninger til udgaver

En stor del af ONP's brugere anvender ordbogen til at slå ord op når de læser en norrøn tekst. For dem der kan lidt norrønt, men stadig har brug for hjælp med sværere ord og betydninger, er ordbogen meget nyttig. På websitets citatsider hentes mange oplysninger om citatet: bl.a. tekstudgaven, håndskriftet, citatsedlen osv. I de fleste tilfælde (99 %) er den scannede udgave uden ophavsret eller dækket af ONP's aftale med CopyDan så siden kan vises sammen med citatet.

Da databasen har henvisninger til specifikke sider og linjer i alle anvendte udgaver, og desuden har billeder af dem, kan dataene bruges den modsatte vej: Programmet kan hente en side fra en udgave og dertil alle citater der er registreret i ordbogen på den pågældende side. Denne funktion kaldes ONP Reader og er tilgængelig gennem den enkelte citatside, samt værkregistret og bibliografien (jf. figur 3).

Gennemsnitligt har ONP excerperet omkring 7 % af alle ord i hele korpusset af norrøne tekster, og 4 % af alle ord i hele korpusset er forsynet med en definition i ONP. Man kan forvente at de redaktører der i sin tid foretog excerperingen, især valgte eksempler på mindre hyppige ord og betydninger, samt dem der vurderedes som gode eksempler. Disse excerperede ord og deres betydninger er derfor dem som læseren plejer at have mest brug for når de



ONP Reader *VidriLS44(1995) (Vidriðá líkams ok sálar)* in Gunnar Harðarson 1995 p. 229 (38)

TEXTUS 229

11 Þv felle rís vpp eigi er sa farin er felle hin er farin er orkaz ma ok vill eigi  
sialfr biarga ser. En hvat of eft verit til vmbota ok opt til meina Rís er  
vpp er þv felle mat er til vmbota. Ef þv bleyðir hiarta þitt sv bleyði  
gæðir þig kemr þv þer til tara þv skola þig. er ok fasta til vmbota beinn  
vokvr mersvongar olmsv gæfir fyrir gefa öðrvm sin misverka vanda  
beinn reitlles er i mistelli hefir þu hefna a þeim illkv er ofþrair erv at  
hallda a sime misti at eigi tynir hanni bæði ser ok öðrvm með illvm  
dæmvm af því at hann er ofþrahaldur a illkv sinni ok er minna hans ton  
eins en magna ok er því hefndin salv botar verk ok synda laeson ef sa  
hefir er gð hefir þat valid i bond grett. Ok er sa reir lagahöfngi er  
hefir eftir logvm ef log erv eigi of frek ok hefir hann fyrir Gvðs sakir  
gövm meinvna til fíðlar ok hefir með avtm en eigi með hefirvm með  
ast við gvð ok fólkt ok sal sialfr sine. Sialfr hitr ok magr abru ficia til  
vmbota. Se hverri mikli málldi vana drotins er hann gerði þig þa er þv  
vart eigi til þerr fagðar ok þv skaldur i hana ríki þat stáð fylia er  
engillin tyndi hann gaf þer at vera vel því at ekki matir þv þer fyrir vnan  
hann þan hann gaf þer erlan at hycia gott ok vilia at vinna ok megin  
Öðv tyndir þv þerr hann galdur þerra cneðv mikva ongm fan þer þer  
anan slikan beina mann hann gefr þer stólin ok vit til at veria enda geti-  
dr þer en með mikv þv at þv tynir með heimkv enda er sia mikvinn æ  
bein hervor off nem þv kettir til ok hervor svöðg ok saorvg þv er svo er  
hann ast mikli við þig. Nv leo moti allt þat gott er þv mat þv er þat  
meogvm lfvim minna hann tilgerðv. Þv þv mat svava ek em faneð  
snoð ok meik ok hefir eigi ofni til framgöva. Sals þv segir eigi sals þv  
mat hafa ast of þv villt hefir þv ok sialla þig þerra lvti kretr hann ok  
kretr sine. En hvat of etni skortu les fram göðvm vilian of hann hefir  
mikil veitt þer ok gerir þv því mikil gott þat er hann vanda. en of hann  
hefir lítit veitt þer þa er þat ok þitt gagn hann vill reyna þig i þinni neiv

124b

L polluta es habes lacrimas quibus iterum te abluas. iterum unctio emaciat  
in te i per benum et pian deosionem iterum te uogis. iterum sciuntis dia-  
tarno confecta es: iterum lacrimis abluta et pie contritionis unctioe  
renouata ad reflectioem tuam redies. Vile quam pia dispensatione tibi  
obsequi conueniat. Non habetur et datum est tibi predicator et restantur  
tibi misquam \*derelinqueris ut scias quantum ille te diligit a quo amaris  
non nisi te pendere et idco tanta patientia expectat et concedit pie negi-  
tior. reuoluit lacrimas iterum unctioe iterum in te aduersa renouat. Di. 124b

1 orkaz = \*orka vb. (146);  
1 rís = \*risa vb. (276); [af e-u etc.] [fyrir e-m] [með e-m]  
1 fellir = \*falla vb. (1660)  
2 hvar = \*hver pron. interrog (985); 1 [e-s / e-u] hvad → 1)  
|| (i retoriske spærgsmål):  
2 vmbota = umbót sb. f. (74)  
3 bleyðir = \*bleyða vb. (16); [e-t] blöðgore, yðmyge i soften, humble  
3 bleyði = bleyði sb. f. (20); 2) hliðhet i sofstness  
4 kemr = \*koma vb. (3440); [e-s]  
4 skola = \*skola vb. (19); [e-n/e-t]  
5 mersvongar = messusvongr sb. m. (33)  
5 misverka = misverki sb. m. (36)  
5 fyrir gefa = fyrirgefa vb. (46); 1 [e-m] [e-t] tilgive (ngn) (ngt), forlade (ngn) (ngt)  
6 reitlæis = réttileiðis adv. (22)  
6 þrair = \*þrár adj. (46); 2) [▪] stædig, trodsig, ubøjelig, egensindig  
6 misfelli = misfelli sb. n. (32)  
6 hitt = hitta vb. (233); [e-t]  
8 ofþrahaldr = ofþráhaldr adj. (1)  
8 þrahaldr = þráhaldr adj. (4)  
9 synda lausn = syndalausn sb. f. (38)  
9 salv botar verk = salubótarverk sb. n. (1)  
10 lagahöfngi = lagahöfngi sb. m. (1)

Figur 3: ONP Reader med billede af udgaven og ONP's citater fra udgavens side med eventuelle definitioner og andre oplysninger <onp.ku.dk/r10680-229> (april 2022).

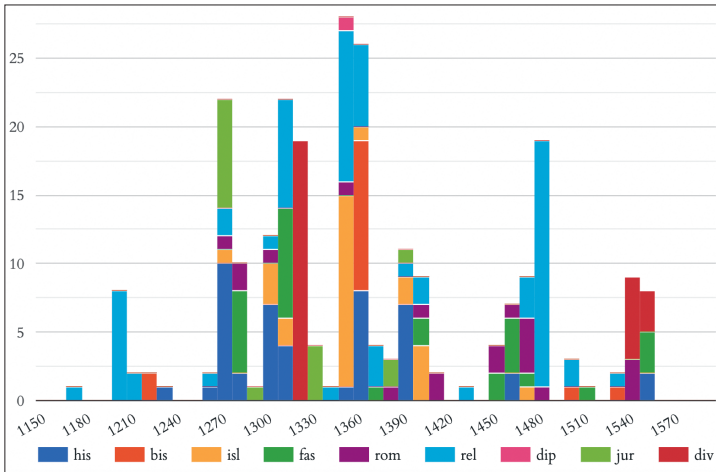
læser en ukendt tekst. Således har gennemsnitligt 4 % af alle ord på en side en definition i ONP på dansk og desuden enten ONP's egen engelske definition eller en automatisk oversættelse af den danske. I de øvrige tilfælde findes der stadig relevante oplysninger for brugeren: opslagsordet, links til andre ordbøger, eventuelle paralleller og grammatiske oplysninger.

## 5.2. Henvisninger til traditionelle materialer

Oplysningerne om udgaverne er et eksempel på at ordbogen kan bruges fra andre indgangsvinkler end ordforrådet alene. Gennem webapplikationen kan man også finde en tekst i en udgave og finde alle citater i ONP der er hentet fra tekstudgaven, og herved få et overblik over tekstens ordforråd.

Ordbogen er designet ud fra det princip at alle citater i ordbogen kan føres tilbage til en fysisk kilde – et håndskrift. ONP har samlet oplysninger om næsten 5.000 håndskrifter og diplomers

datering samt deres indhold. Man kan derved bruge ONP som et katalog over den norrøne prosalitteraturs kilder, og under hvert håndskrift opstilles alle citater der stammer fra det pågældende håndskrift, hvilket kan bruges til en undersøgelse af den fysiske kildes sprogbrug. For hvert lemma kan man få et overblik over ordets diakroniske forekomst i håndskrifterne, se figur 4.



Figur 4: Citater under *drekka* ('drikke') efter håndskriftsdato og genre <onp.ku.dk/015159> (april 2022).

Fra citatsiderne linkes der også til to andre ressourcer der har billeder af håndskrifterne: Handrit.is, som har digitale billeder af mange norrøne håndskrifter, og NorS Sprogsamlinger, som offentliggør scannede fotografier af håndskrifter i Den Arnamagnæanske Samlings fotosamlinger. Brugeren kan derfor få adgang til billeder af det oprindelige fysiske kildemateriale gennem websitet.

Som en del af redaktionsprocessen har ordbogen sammenlignet citaterne med udenlandske kilder og andre relevante paralleller til ordbogens kildeværker. I 35.000 citater findes der henvisninger til disse kilder som også er linket til registrene. Man kan så finde parallellerne under hvert værk i registrene, og webapplika-

tionen kan generere en liste over alle citater hvor den parallelle tekst er blevet registreret med tilsvarende ord. Parallellerne findes også under ordbogsartiklerne selv, men den konsekvente måde hvorpå de er blevet registreret, betyder at ordforrådet også kan undersøges med de udenlandske kilder som udgangspunkt.

### 5.3. Forbindelse til digitale korpusser

Selvom ONP er en traditionel, excerperingsbaseret ordbog, har den flere fælles referencepunkter med nyere digitale korpusser. ONP's lemmaliste bruges som standarden for lemmatisering af digitale tekster, f.eks. i The Skaldic Project og Menota (se Haugen 2019, §11.2). De to projekter har stor nytte af ONP fordi der er tale om håndskriftsbaserede udgaver af norrøne tekster. The Skaldic Project er i gang med at redigere en stor del af det norrøne poesi-korpus fra håndskrifterne, og det inkluderer også en oversættelse til engelsk med links på ordniveau mellem kildeteksten og oversættelsen. Menota arkiverer og offentliggør TEI-baserede udgaver af enkelte norrøne (og andre nordiske) håndskrifter. Disse udgaver har derudover ofte oplysninger om ordenes morfologi.

I The Skaldic Project's tilfælde har ONP givet adgang til sin digitale lemmaliste, og projektet har brugt den til at lemmatisere korpusset. De to projekter kan linkes sammen entydigt med hensyn til homografer og ortografi. ONP dækker ikke poetiske ord og tekster, men citaterne i ONP kan suppleres med ord fra The Skaldic Project's korpus (jf. ovenfor i afsnit 3.1).

Menotas tekstarkiv indeholder især prosalitteratur. Udgaverne bruger som regel ONP's ordliste til at identificere deres lemmer, men uden præcisering af homografer. Menotas tekster har et andet fælles referencepunkt med ONP, da udgaverne er baseret direkte på enkelte håndskrifter. I princippet kan ord fra Menota linkes ikke alene med ONP's lemmaliste, men også med citaterne selv. ONP anvender nogle af Menota-teksterne så lemmerne

KØBENHAVNS UNIVERSITET

ONP

Hjem

◀

◄

Holm p  
Barlaam

(Rindal

4) (fig.)

Sem b

sinum

net sit

Manuscript page (Menota)

Holm perg 6 fol — *ed. Magnus Rindal*

26r/a1 Oc allzskonar sinn vilia til hanns ventt

26r/a2 þeim visar hann til heimilis firir sina

26r/a3 starflaun til pinsla oc kuala oc til

26r/a4 allzskonar illgiætess oc ber enga sorg

26r/a5 firir þo at hann se ve

26r/a6 oc fysizt æ til þess at

26r/a7 fae vfagnað af hans f

26r/a8 hann hevir eina svik

26r/a9 þa reisir hann **net** <

26r/a10 at veiða þa er hans

26r/a11 Slikan mala oc starf

26r/a12 þeir vpp er slikum

26r/a13 þiona er bæðe er ill

26r/a14 oc þo margsløgr fa þeir slikt

26r/a15 skippti a er sik firra goðo raðe

**net** *sb. n.* [64] *BarLA 47<sup>24</sup>*

≠ *lat.* Illos enim male seduxit et retibus  
conclisit, ad istos uero rursus artem suam  
transferre conatur *JDamBarl 41<sup>9-10</sup>*

4) (fig.) (om ondskabens/djavlens  
følder/fristelser)

Figur 5: Side fra Menotas korpus, med links til ONP's citater, definitioner m.m.

kan linkes utvetydigt til ONP's ordliste, og korpusset kan supplere ONP's citater med alle eksempler fra enkelte tekster (jf. figur 5). Yderligere automatisk og manuel bearbejdning kan linke selve citaterne til Menotas ord i fuld kontekst (se Wills, Jóhannsson & Battista 2018). I disse tilfælde er det muligt at generere en interaktiv tekst hvor ONP's oplysninger om de excerperede citater kan hentes direkte ved at klikke på ordene i Menotas udgave.

#### 5.4. Digital adgang til ONP's data

ONP's webapplikation bruger flere eksterne ressourcer der hentes via en Application Programming Interface (API), som nævnt ovenfor (se afsnit 3.1). ONP gør også allerede nu sit eget materiale tilgængeligt for eksterne applikationer på en lignende måde, og den bruger således de relevante standarder.

ELEXIS-projektet arbejder på forskellige områder på at bygge en digital leksikografisk infrastruktur. Projektet har udviklet en standard-API for digital adgang til ordbøger (jf. ELEXIS Protocol for accessing dictionaries (1.1)), som specificerer hvordan eksterne programmer kan hente oplysninger om en ordbog, søge i lemmalisten og hente artikler. ELEXIS' API er blevet implementeret i ONP. Et eksternt program kan derfor søge i ONP's ordliste, hente hele ordlisten, finde opslagsord ved hjælp af en søgestreng (for eksempel efter første eller sidste del af ord) og hente digitale versioner af enkelte ordbogsartikler.

ELEXIS' API-specifikationer giver mulighed for at bruge to standarder til ordbogsartiklerne: Ontolex-Lemon (Bosque-Gil & Gracia 2019) i to forskellige digitale formater, eller TEI. Ontolex er ikke ideelt til historiske ordbøger da det er udarbejdet for moderne ordbøger og ikke specificerer kodning af henvisninger til tekster. Vi implementerer det alligevel da det kan bruges mere konsekvent end TEI.

I princippet kan API'en bruges til at integrere ONP's materiale i andre portaler og applikationer. ONP er i gang med at samarbejde med *málið.is* om at inkludere ONP i søgeresultaterne i deres portal.

## 6. Resultater

På trods af at ONP ikke er færdigredigeret, fungerer den allerede nu som et vigtigt hjælpemiddel for undersøgelser af det norrøne prosasprog og dets kilder. ONP og de digitale ressourcer som er tilgængelige på ONP's hjemmeside, er resultatet af 40 års arbejde med at supplere og effektivisere redigeringsprocessen ved til staidighed at tilføje nye data og gøre dem tilgængelige digitalt. Disse data indeholder righoldige oplysninger om det norrøne sprog, de fysiske og tekstuelle kilder, udenlandske paralleller og ældre ord-

bøger. Den seneste, aktuelle proces har gjort disse ressourcer, både ældre og nyere, tilgængelige for nye brugergrupper der nu kan drage nytte af dem i forskellige anvendelsessituationer. Samtidig bidrager ressourcerne fortsat til at forbedre redigeringsprocessen: Som nævnt ovenfor har de nye funktioner i webapplikationen gjort dele af redaktionsarbejdet væsentligt hurtigere.

ONP trækker også på forskellige eksterne digitale materialer og kilder for at supplere og berige ONP's materiale, som stadig er under udarbejdelse. Disse eksterne ressourcer består af digitale korpuser, digitaliserede ordbøger og andre leksikalske hjælpemidler. Korpuserne bidrager med konkordanser og i visse tilfælde oversættelser til ordene, og de leksikalske ressourcer samt ordbøger bidrager med artikler som fortsat mangler i ONP, målsprog/definitioner (især engelsk), nyere kildemateriale og lingvistiske oplysninger (etymologi og bøjningsoplysninger). Disse processer og ressourcer betyder at brugere altid kan tage udgangspunkt i ONP's hjemmeside for at undersøge et norrønt ord og dets brug, også i de tilfælde hvor ordet endnu ikke er blevet redigeret.

Som et resultat af denne udvikling kan vi konstatere at ONP's webbrugerbase ifølge Google Analytics er vokset fra ca. 1.000 pr. måned (2017) til over 5.000 pr. måned (2022). Brugerne besøger flere sider og bruger flere funktioner end tidligere. Dog bruges de mange muligheder der ligger dybere inde i webapplikationen, ikke i samme udstrækning. Vi arbejder fortsat på at gøre dem mere synlige og tilgængelige for vores brugere.

De metoder der beskrives og fremlægges i denne artikel, viser hvordan en digital ordbog kan sammenkoble funktioner fra forskellige ressourcer for at supplere sit eget materiale. Ordbogen – eller den hjemmeside hvor ordbogen publiceres – kan derfor altid fungere som udgangspunkt i en undersøgelse af sprogets, i dette tilfælde det norrøne prosasprogs, ordforråd.

## Litteratur

## Ordbøger

*Blöndal* = Blöndal, Sigfús (1920-1924): *Íslandsk-dansk Ordbog. Íslensk-dönsk orðabók*. Reykjavík: Verslun Þórarins B. Þorlákssonar / Aschehoug.

*CIV* = Cleasby, Richard, Gudbrand Vigfusson [Guðbrandur Vigfússon] and W. A. Craigie (1957): *An Icelandic-English Dictionary*. 2. udg. Oxford: Clarendon.

*Fritzner* = Fritzner, Johan (1883-96): *Ordbog over det gamle norske sprog*. Kristiania (Oslo): Den norske forlagsforening.

*Hertzberg* = Hertzberg, Ebbe (1895): *Glossarium. I: Storm, Gustav og Hertzberg, Ebbe (udg.): Norges gamle Love indtil 1387*. Bind 5. Kristiania.

*LP* = Finnur Jónsson (1931): *Lexicon poeticum antiquæ linguæ septentrionalis: Ordbog over det norsk-islandske skjaldesprog oprindelig forfattet af Sveinbjörn Egilsson*. 2. udg. Copenhagen: Møller.

*málið.is* = <malid.is> (april 2022).

*MED* = Robert E. Lewis, et al. (1952-2001): *Middle English Dictionary*. Ann Arbor: University of Michigan Press. Onlineudgave: Frances McSparran, et al. (2000-2018): *Middle English Compendium*. Ann Arbor: University of Michigan Library. <quod.lib.umich.edu/m/middle-english-dictionary> (april 2022).

*ONP* = Degnbol, Helle, Bent Chr. Jacobsen, Eva Rode, Christopher Sanders, Þorbjörg Helgadóttir, James E. Knirk, Aldís Sigurðardóttir, Alex Speed Kjeldsen, Ellert Þór Jóhannsson, Pernille Ellyton, Johnny Lindholm & Tarrin Wills (1989-): *Ordbog over det norrøne prosasprog // A Dictionary of Old Norse Prose*. København: Den Arnamagnæanske Kommission. <onp.ku.dk> (april 2022).

## Anden litteratur

- Battista, Simonetta & Ellert Þór Jóhannsson (2014): Ordbog over det norrøne prosasprog 2004-2014: fra trykt udgave til netversion. I: *LEDA-nyt* 57, 6-23.
- Bosque-Gil, Julia & Jorge Gracia (udg.) (2019): *The OntoLex Lemon Lexicography Module*. <w3.org/2019/09/lexicog/> (april 2022).
- ELEXIS: European Lexicographic Infrastructure. <elex.is> (april 2022).
- Haugen, Odd Einar (2019): Linguistic Annotation. I: Odd Einar Haugen (udg.): *The Menota handbook: Guidelines for the electronic encoding of Medieval Nordic primary sources*. Version 3.0. Bergen: Medieval Nordic Text Archive. <menota.org/HB3\_ch11.xml>.
- Jóhannsson, Ellert Þór & Simonetta Battista (2016): Editing and Presenting Complex Source Material in an Online Dictionary: The Case of ONP. I: Margalitadze, Tinatin et al. (udg.): *Proceedings of the XVII EURALEX International Congress*. Tbilisi: Tbilisi University Press, 117-128.
- Mugglestone, Lynda (2005): *Lost for Words: The Hidden History of the Oxford English Dictionary*. Yale: Yale University Press.
- TEI Consortium (2022): *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. 4.4. TEI Consortium. <tei-c.org/Guidelines/P5/> (april 2022).
- Wills, Tarrin, Ellert Þór Jóhannsson & Simonetta Battista (2018): Linking Corpus Data to an Excerpt-based Historical Dictionary. I: Čibej, Jaka et. al. (udg.): *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*, 979-988.
- Wills, Tarrin, Simonetta Battista & Ellert Þór Jóhannsson (2021): Making an ongoing historical dictionary accessible to an English-speaking audience. Euralex XIX, Greece. <euralex>



[lex2020.gr/wp-content/uploads/2021/10/EURALEX2020\\_BookOfAbstracts-Preview.pdf](http://lex2020.gr/wp-content/uploads/2021/10/EURALEX2020_BookOfAbstracts-Preview.pdf)> (april 2022).

## Anvendte korpuser

Handrit.is = Handrit.org / Handrit.is. <[handrit.is](http://handrit.is)> (april 2022).

Menota = Medieval Nordic Text Archive: Public Catalogue. <[clarino.uib.no/menota/catalogue](http://clarino.uib.no/menota/catalogue)> (april 2022).

NorS Sprogsamlinger = Institut for Nordiske Studier og Sprogvidenskabs Digitale Sprogsamlinger. <[sprogsamlinger.ku.dk](http://sprogsamlinger.ku.dk)> (august 2022).

The Skaldic Project = Skaldic Poetry of the Scandinavian Middle Ages. <[skaldic.org](http://skaldic.org)> (april 2022).

Tarrin Wills  
Ordbogsredaktør og -leder  
Ordbog over det norrøne prosasprog  
Københavns Universitet  
Njalsgade 136  
DK-2300 København S  
[tarrin@hum.ku.dk](mailto:tarrin@hum.ku.dk)