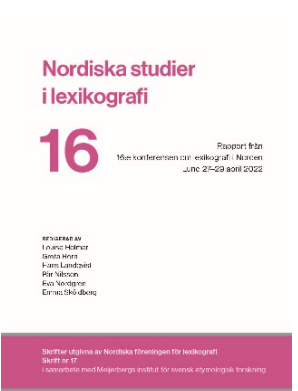


# NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Hvordan holde trit med tiden i en historisk ordbok som også beskriver samtiden?	
Forfatter:	Carina Nilstun	
Kilde:	Nordiska studier i lexicografi 16, 2023, s. 257–265	
URL:	<a href="https://tidsskrift.dk/nsil/issue/archive">https://tidsskrift.dk/nsil/issue/archive</a>	

© 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.

# Hvordan holde tritt med tiden i en historisk ordbok som også beskriver samtiden?

*Carina Nilstun*

*The Norwegian Academy Dictionary* (NAOB) is a historical dictionary, describing the variety bokmål/riksmål from early 1800 until today. It is also a literary dictionary, with entries built upon the actual language use in texts. Furthermore, NAOB is a contemporary dictionary. The entry head reflects today's spelling, inflection and pronunciation, and the editorial examples belong to the present. NAOB is in its origin a reading comprehension dictionary but can today also serve as a production dictionary. NAOB is typologically wide, and to fully be a contemporary dictionary it must receive continuous refills of current material on every level: words, senses and fixed expressions, as well as quotations. Quotations from the recent years should be rich in quantity, given that the base consists of more than 200 000 entries and more than 300 000 quotations. Publications from after year 2000 are not publicly available in Norway. But these 22 years are of immense importance for our contemporary dictionary. In this paper I will show how the lexicographers of NAOB work their way around this.

NØKKELORD: praktisk leksikografi, litterær ordbok, leksikografisk metode

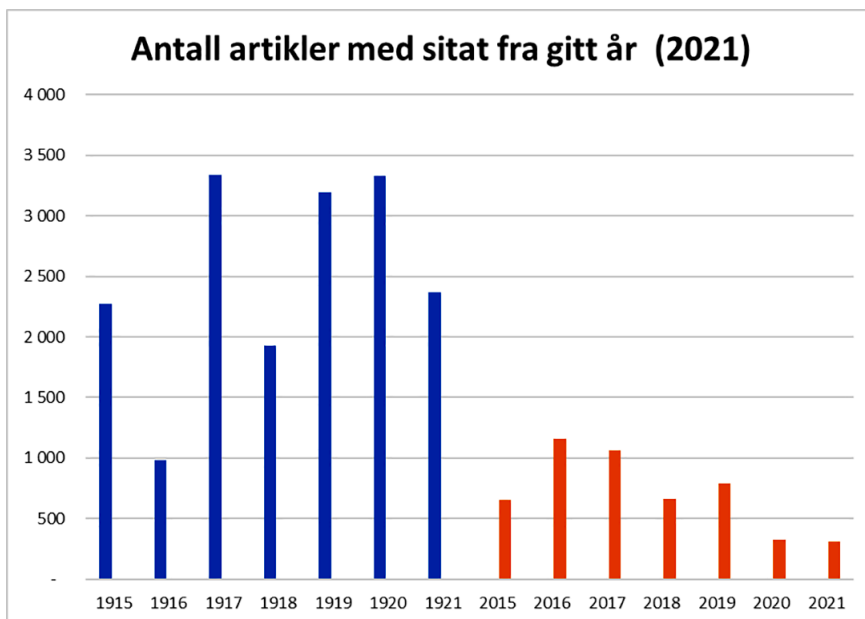
## 1. Innledning

Det Norske Akademis ordbok (NAOB; <https://naob.no>) er en historisk ordbok som beskriver bokmålet og riksmålet fra tidlig 1800-tall og frem til i dag. NAOB er også en litterær ordbok, slik at artiklene er bygget opp med utgangspunkt i den bruk som finnes i skriftlige kilder. Dessuten er NAOB en samtidsordbok – artiklenes hodeopplysninger følger dagens språk (med tanke på staving, bøyning og uttale), og de redaksjonelle eksemplene tilhører alltid nåtiden. Skal NAOB fungere som samtidsordbok, må vi sørge for kontinuerlig påfyll av nytt ordstoff på alle nivåer: ord, betydninger og faste uttrykk. Videre må sitater fra nåtiden tas inn uten særlig forsinkelse og i et omfang som monner i en base med over 200 000 oppslagsord og mer enn 300 000 sitater.

NAOB finansieres dels av staten ved Kultur- og likestillingsdepartementet, dels av private stiftelser. For 2022 og 2023 har prosjektet fått til sammen 800 000 kroner fra stiftelsen Fritt Ord (<https://frittord.no/>) for å øke andelen sakprosa i NAOB og å øke andelen litteratur fra de aller siste årene. Redaksjonen består av tre leksikografer, en mindre del av den ene stillingen går til ledelse og administrasjon.

## 2. Endret siktemål

Når vi sier at NAOB beskriver norsk bokmål fra 1800-tallet og frem til i dag, er i dag bokstavelig ment: høyre ende av tidsaksen er åpen. NAOB er dokumenterende og litterær; vi dokumenterer ordforrådet gjennom den faktiske bruken i skrift. Ordbokens kildegrunnlag er åpent og utvides kontinuerlig. Samtiden skal være til stede på alle ordbokens nivåer: i lemmautvalget, i betydningsstrukturen og i kildene det siteres fra. Vårt forelegg er *Norsk Riksmålsordbok*, som har dokumentert store forfattere som blant andre de to Henrikene Wergeland og Ibsen svært grundig. Så var det også primært en resepsjonsordbok *Norsk Riksmålsordbok* var. NAOB er også en resepsjonsordbok, men vi har i større grad enn våre forgjengere et mål om at NAOB skal speile *språket*, og at samtiden hele tiden skal være present i ordboken. Den enkelte forfatter spiller en mindre rolle i NAOB, mens man i *Norsk Riksmålsordbok* dokumenterte enkelte forfatterskap nærmest uttømmende. Nå er det entydig språket som er objektet som dokumenteres, og det gjøres ved hjelp av litterære sitater. Ingen moderne forfatter er så grundig representert som tilfelle er med Wergeland og Ibsen. Med en så grundig dokumentasjon av den eldre litteraturen kreves det en målrettet innsats for å synliggjøre samtiden.



FIGUR 1. Antall artikler i NAOB med sitater fra årene 1915–1921 og 2015–2021. Tallene er hentet i 2021.

Figur 1 viser hvordan antall artikler med sitater fra årene 1915–1921 er langt høyere enn tallene hundre år senere. Tallene er fra før det nye prosjektet med å øke antall sitater fra de siste årene ble påbegynt. Det er ikke noe mål at alle år skal være likt representert, og heller ikke at årstallene skal «fylles opp» med det samme, for vi legger stadig til sitater også fra eldre litteratur. Men antall sitater fra senere år bør noe raskere kunne bli en del flere.

Ved inngangen til 2022 hadde ordboken 5000 sitater fra tidsrommet 2015–2021. Ved årets slutt var det 7700 sitater fra tidsrommet 2015–2022. Høyst er antallet for 2021, hvor det per mars 2023 er over 2000 sitater. Endringene kan leseren selv finne frem til ved å fritekstsøke årstallet man vil undersøke på [naob.no](http://naob.no).

## 2.1. Ulik bruk

Hvis en bruker søker opp ord av typen *stjerneblomme* og *begerbluss* vil hen naturlig nok møte sitater fra fortiden, og bare det. Søkes det på ord ord som *snitche* og *tæsje*, er det moderne kilder som er sitert. Sånn må

det være. Men søker man opp ord fra kjerneordforrådet, skal aktualiteten fremgå gjennom litteraturen som er sitert. Som samtidfokusert NAOB-bruker skal man gjennom bruk alene kunne vite at NAOBs korpus ikke er lukket, ved at man stadig treffer på nyere årstall i artiklenes sitatseksjon. Vi har ingen ambisjoner om å finne det eldste belegget for hvert ord i NAOB, men vi skal hele tiden ha ordbokens tidsspenn med oss i redigeringen. Ordet *fullvaksinere* kom inn i NAOB i 2021, men det viste seg at det ikke var nytt. Et eldre sitat kom derfor også med. Som oftest er det motsatt: Vi har artikler som har vært med helt fra *Norsk Riksmålsordbok*, med eldre sitater. Da er det viktig at nyeste sitat for et fullt brukbart ord ikke er fra for eksempel 1960-tallet. Det er utenkelig å komme igjennom hele basen i løpet av få år, men gjennom Fritt Ord-prosjektet vil vi få sitatfornyte artiklene for en stor mengde aktuelle ord.

### 3. Kildegrunlaget og basene det aksesseres fra

I Norge ligger det meste av bøker utgitt frem til år 2000 digitalt og fritt tilgjengelig hos Nasjonalbiblioteket ([www.nb.no](http://www.nb.no)), men hva så med de siste 22 årene? Vi kan ikke nøye oss med sitater fra aviser og tidsskrifter når vi skal vise språket i bruk de to siste tiårene. Det skapes lett et inntrykk av at NAOB er tilbakeskuende, med den store mengden sitater fra Ibsen og andre 1800-tallsforfattere vi har med oss fra *Norsk Riksmålsordbok*. NAOB-redaksjonen ønsker i de kommende årene å legge enda større vekt på å få inn sitateksempler fra ny litteratur tidlig. I det følgende vil jeg vise hvordan redaksjonen arbeider for å ha på plass et representativt utvalg av ny litteratur allerede i utgivelsesåret og vise hvordan vi har lagt opp dette arbeidet for å få mest mulig samtidsordbok for pengene.

NAOBs litteraturliste består av godt over 6000 bøker og rundt 800 aviser og tidsskrifter (<https://naob.no/litteratur>). Vi har ikke noe korpus som rommer dem alle, så hvordan når vi våre kilder?

For aviser og tidsskrifter bruker vi A-tekst – et abonnementsbasert mediearkiv med over 300 aviser og tidsskrifter, eller vi kan også gå rett til kilden, til avisens nettutgave. For sakprosa og skjønnlitteratur har vi flere gode baser, men de fleste har et litteraturutvalg som stopper i år 2000, noen også før. Litteraturen fra de manglende 22 årene kan vi som dokumenterende ordbok ikke unnvære. Vi skal se nærmere på to av basene, og også se på hvordan vi får tak i den nyeste litteraturen.

### 3.1. Nettbiblioteket

Nettbiblioteket (tidligere kalt Bokhylla) er en digital tjeneste hos Nasjonalbiblioteket som tilbyr søk i norske bokutgivelser, fritt tilgjengelig. Nettbiblioteket er en språkskatt og en ressurs vi bruker mye. Fordelene er mange: Det enorme omfanget er svært nyttig siden vi også beskriver forholdsvise sjeldne ord, aviser og fagtidsskrifter samlet på ett sted effektiviserer arbeidet og nettstedet har mange funksjoner, eksempelvis N-gram-søk (<https://www.nb.no/ngram/>). Haken ved Nettbiblioteket er at det som er fritt tilgjengelig for allmennheten, er utgivelser som er fra år 2000 eller eldre, med unntak av noen verk som rettighetshavere har holdt tilbake. Stoff fra etter år 2001 er kun tilgjengelig med en særskilt tilgang som bare tildeles utdanningsinstitusjoner. NAOB, som ikke er tilknyttet en utdanningsinstitusjon, får det ikke. Men det finnes andre gode løsninger.

### 3.2. DocFetcher

DocFetcher er en ”open source desktop search application”, altså programvare som lar deg søke i en mengde av filer og returnerer filene som inneholder søkestrengen (<http://docfetcher.sourceforge.net/en/index.html>). I praksis blir dette et korpus. Søkemessig er det ganske likt som Nettbiblioteket: Man kan trunkere og gjøre flerordssøk, men man får ikke opp noen konkordans. Vi har altså liten nytte av å fylle DocFetcher med filer som allerede er fritt tilgjengelige, for dem når vi gjennom Nettbiblioteket. Det vi har behov for, er filene med utgivelser fra 2001 og senere. Vi har derfor kontaktet forlagene direkte, fortalt dem om vårt behov og bedt om å få tilsendt utvalgte PDF-filer. Disse filene legges inn i DocFetcher. Vi velger ut forfattere og bøker som preger samtiden, og vi velger ut sakprosa, både ut fra bøkernes utbredelse og ut fra fag/tema. Vi vektlegger også bevisst en balanse mellom mannlige og kvinnelige forfattere. Forlagene er velvillige og sender oss de filene vi ber om. Både forfattere og forlag bidrar med glede til vårt prosjekt; de ser viktigheten av dokumentasjons- og formidlingsarbeidet. Vi har fått på plass en standardkontrakt, slik at forlagene er trygge på at vi ikke videredistribuerer filene.

For vårt arbeid er det en ulempe at Nettbiblioteket ikke kan skille ut oversatt litteratur (som ikke er del av NAOBs tekstgrunnlag). Med det skreddersydde DocFetcher-korpuset er det problemet løst. Denne avgren-

singen gjør også at hver kilde blir mer brukt. DocFetcher har videre den fordel at vi kan navngi filene og utnytte dette i sorteringen av treffene. For bøker vi ønsker å sitere ekstra mye fra, kan filnavnene starte med 01. Disse filene legger seg da øverst når trefflisten sorteres på filnavn. DocFetcher-korpuset brukes i den generelle redigeringen, men også mer målrettet. Den generelle redigeringen vil si når vi arbeider med en ny artikkel, eller arbeider med å få inn nye sitater i eksisterende artikler. Da kan vi bruke kilder med prefiks 01 i navnet for å styrke andelen sitater fra nyere litteratur eller sakprosasitater. Målrettet redigering er det f.eks. når vi i fagspråkprosjektet skal styrke konkrete fag eller domener. Da kan vi sanke både ord og sitater fra DocFetcher-korpuset.

Med DocFetcher får vi altså arbeidet svært effektivt med begge Fritt Ord-prosjektene samtidig; ny litteratur finner veien inn i NAOB, både i form av sakprosa og skjønnlitteratur. Ordforrådet kan finnes i NAOB fra før og bare suppleres med sitat, nytt ordforråd, nye betydninger eller nye faste uttrykk kan legges til.

## 4. Metodikk for sakprosaprosjektet

### 4.1. Ekserpering

Ekserpering er en svært nyttig metode for den type leksikografisk arbeid som utføres i NAOB, men det er en tidkrevende arbeidsmåte. Vi har såpass begrensede ressurser at vi ikke kan sitte og lese hele bøker i arbeidstiden. En mulighet er å streke under godt ordstoff i bøker man allikevel leser på fritiden, og så bearbeide fangsten i arbeidstiden. Dette er et svært nyttig tilskudd til ordboken, men det er på ingen måte forventet av leksikografene, og dermed ikke en metodikk vi kan basere ordbokens utvikling på. Utvalget blir dessuten tilfeldig, og dette er altså bare et tilskudd til det mer planmessige arbeidet.

Boka *Dyrespor* av Arnodd Håpnæs fra 2021 tar for seg dyrespor og spor tegn for alle ville dyr som finnes i Norge. Den favner felter som zoologi, biologi og økologi, og den er skrevet for allmennheten. Den type bøker er gode kilder for oss ved at de gir gode, informative sitater og viktig ordforråd. Det samme gjelder eksempelvis også pubertetsbøkene *Jenteboka* (Brochmann & Dahl 2019) og *Gutteboka* (Brochmann & Dahl 2021), som særlig dekker medisin og anatomi. I tillegg er det naturligvis alltid også mye allmennspråklig stoff i sakprosa som også siteres. Mye

sakprosa er det enklere å høste fra enn skjønnlitteratur, siden sakprosautgivelser ofte har en innholdsfortegnelse og kanskje også en ordliste. Dermed har man tilgang til sentrale ord fra faget eller domenet. Mange sitater kan også gjenbrukes. Et sitat som «fjellreven ... beveger seg [som regel] i lett galopp, mens rødreven oftest traver» (Håpnes 2021:82) er like godt som belegg i alle de fire artiklene *rødrev*, *galopp*, *trave* og *fjellrev*. Tilsvarende har vi nylig tatt inn i litteraturlisten en trilogi om funksjonshemning, skrevet av Jan Grue. Her er det ingen innholdsfortegnelse eller ordliste, men vi kan søke opp domenerrelevant stoff, for eksempel sammensetninger med *rullestol*. På den måten fikk *rullestolrampe*, *rullestoltilgjengelig* og *rullestoltilgjengelighet* en plass i NAOB. Dette igjen inngår i et annet av 2022s prosjekter, nemlig mangfold i ordboken. Prosjektet er tematisert i Hanne Lauvstads artikkel i denne konferanserapporten (Lauvstad 2023). Med PDF-filene har vi hatt god nytte av å foreta ekserpering ved målrettede søk ut fra bokens tema, ved gjenbruk av sitater og gjennom skumlesing, og dessuten ved bruk av språkteknologisk hjelpemidler.

## 4.2. Verktøyet Sketch Engine

Det mer manuelle arbeidet har altså fortsatt mye for seg i NAOB-redigeringen, men språkteknologiske hjelpemidler er selvsagt også viktige. Med PDF-filene for hånden, er Sketch Engine et ypperlig verktøy (<https://www.sketchengine.eu/>). Der kompilerer man egne korpus på nærmest ingen tid, fra filer man selv besitter. I sakprosa-prosjektet er det nyttig å lage korpus av én og én bok, eller av flere bøker som tilhører samme fag eller område. Disse sammenlignes så mot et mye større korpus for å skille ut tekstens nøkkelord, det vil si de ordene som preger teksten ved å ha høyere frekvens enn i sammenligningskorpuset. Det man ikke får høstet rikt av på denne måten, er allmennspråket i sakprosaen, fagrelevant stoff uten høy frekvens eller syntaktiske konstruksjoner, som burde vises frem i ordboken. Der kan skumlesing gi en viss fangst, og også Sketch Engines kollokasjons-funksjon. Sitatet «ulvemotstanderne har en til dels sterkt utviklet motvilje mot forskere, som ofte oppfattes som naturvernernes, forvaltningens og fagstatsrådets forlengede arm» (Rossavik 2021:69), som på en fin måte viser uttrykket *noens forlengede arm*, vil ikke tre frem i en word sketch-analyse, mens leksikografblikket straks vil gjenkjenne sitatets illustrative kvaliteter.



### 4.3. Utvelgelse av litteratur

Redaksjonen arbeider i fellesskap med å plukke ut ny litteratur. For skjønnlitteratur vektlegger vi bøker som har fått en stor leserkrets og/eller viktige priser, og således har preget tiden. Eksempler er Linn Ullmanns *Jente*, 1983 (Ullmann 2021), Trude Marsteins *Egne barn* (Marstein 2022), Jan Kjærstads *En tid for å leve* (Kjærstad 2021). På den annen side gjør vi også en innsats for å løfte frem mindre etablerte navn og sjangere som får mindre oppmerksomhet. Barne- og ungdomslitteratur, både fra samtiden og eldre, ønsker vi generelt å dekke bedre. Sakprosa velges både ut fra fagområder hvor NAOB trenger styrking, hvor det har skjedd mye nytt og ut fra hva som preger allmennheten og dermed allmennspråket.

### 5. Avslutning

NAOB skal være oppdatert når det gjelder både ordforrådet og litteraturen, og dette henger sammen. I sakprosa-prosjektet skal vi styrke ordforrådet og mengden sitert litteratur generelt, i tillegg til at vi har sett oss ut noen områder som er ekstra aktuelle gjennom sin plass i allmennheten, blant annet økologi og medisin. Dette er fagområder som trenger mer og mer inn i allmennspråket, og hvor det er behov for gode leksikografiske beskrivelser, slik man kan forvente å finne det i en samtidsordbok. Samtidig vokser bekymringen for norsk som fagspråk fordi engelsk blir stadig mer dominerende. Fagmiljøene selv må bruke norsk og gjennom bruken opprettholde og utvikle norsk fagspråk. En dokumentasjonsordbok som NAOB er en viktig støttespiller i dette arbeidet. Fagspråk er pekt ut som satsningsområde for ordboken i de nærmeste årene, og her er samtiden en forutsetning.

### Litteratur

#### *Ordbøker*

NAOB = *Det Norske Akademis ordbok*. Oslo: Det Norske Akademi for Språk og Litteratur. <<http://naob.no>> (mars 2023).

*Annen litteratur*

Lauvstad, Hanne (2023): Sensitive ord i *Det Norske Akademis ordbok*. Utfordringer i diakron leksikografi. I: Holmer, Louise et al. (red.), *Nordiska studier i leksikografi* 16. Lund & Göteborg: Nordiska föreningen för leksikografi, 213–224.

*Sitert og omtalt kildelitteratur*

Brochmann, Nina og Dahl, Ellen Støkken (2021): *Gutteboka. Din guide til puberteten*. Oslo: Aschehoug.

Brochmann, Nina og Dahl, Ellen Støkken (2019): *Jenteboka. Ellen og Ninas guide til puberteten*. Oslo: Aschehoug.

Grue, Jan (2018): *Jeg lever et liv som ligner deres: En levnetsbeskrivelse*. Oslo: Gyldendal.

Grue, Jan (2021): *Hvis jeg faller: En beretning om usynlig arbeid*. Oslo: Gyldendal.

Grue, Jan (2022): *Prøve og feile*. Oslo: Gyldendal.

Håpnes, Arnodd (2021): *Dyrespor. Og andre tegn etter dyr*. Oslo: J.M. Stenersens forlag A.S.

Kjærstad, Jan (2021): *En tid for å leve*. Oslo: Aschehoug.

Marstein, Trude (2022): *Egne barn*. Oslo: Gyldendal.

Rossavik, Frank (2021): *Ulv? Ulv! En bok om rovdyr og mennesker i Norge*. Oslo: Cappelen Damm.

Ullmann, Linn (2021): *Jente, 1983*. Oslo: Forlaget Oktober.