



Digital humaniora, mediehistorie og litterære subjektivitetsuttrykk. Om forholdet mellom norsk litteratur og utviklingen av den kommersielle pressen 1855–1900 i et DH-perspektiv

Digital Humanities, Media History, and Literary Expressions of Subjectivity. On the Relationship Between Norwegian Literature and the Development of the Commercial Press 1855–1900 in a DH Perspective

Frode Boasson

Førsteamanuensis i nordisk litteraturvitenskap, NTNU

(f. 1983) Siste publikasjon: «Pleasure Is Not Fun: The Locus Amoenus in Knut Hamsun» i *Mutating Idylls. Uses and Misuses of the Locus Amoenus in European Literature, 1850–1930*. Peter Lang Publishing Group 2019.

frode.boasson@ntnu.no

Anders Skare Malvik

Førsteamanuensis i nordisk litteraturvitenskap, NTNU

(f. 1981) Siste publikasjon: «Mediekritisk ungdomslitteratur.» i Ann Sylvi Larsen (red.), *Norsk boka 5–10*, Oslo: Universitetsforlaget 2018.

anders.malvik@ntnu.no

Abstract

The National Library of Norway (NB) has digitized its entire collection of books. This article examines how to make use of this collection in literary studies. The article focuses on the literature of the late 1800s and asks in particular how the modern press made an impact on Norwegian literature in the late 1800s. Based on a series of close readings of canonical texts, we hypothesise that the rise of the commercial newspaper in Norway initiated a new and ambivalent notion of subjectivity that surfaced in literary texts throughout the second half of the nineteenth century. By way of keyword searches, and digital N-gram and collocation analyses in the NB's digitized archive, we find new empirical material that supports this hypothesis and invites further studies.

Keywords

Digital Humanities, Digitization, Macro Analysis, N-gram, Collocation-analysis, Literary History

1. februar 2018 markerer datoen da Norge som første land i verden gjorde hele sin nasjonale litteraturarv tilgjengelig i digital form. Over 250.000 av Nasjonalbibliotekets (NBs) bøker utgitt før år 2000 er digitalisert og åpent tilgjengelige for allmennheten (Bok365 «Alt på

plass på Bokhylla» 2018)¹. Når hele boksamlingen er digital, er ikke bare bøkene et taste-klikk unna. Hele litteraturhistorien kan i prinsippet analyseres av datamaskiner. Slik sett er den digitale bokhylla en gavepakke til alle som er interessert i norsk litteratur- og bokhistorie. Spørsmålet er bare hvordan vi skal nyttiggjøre oss digitaliseringen: Hvordan skal vi finne frem til og skille ut relevant informasjon? Hvordan skal vi analysere og forstå det vi finner? Ikke minst: Hvor pålitelig er måten vi henter ut informasjon på og hvilken status har den egentlig?

Datamaskiner *leser* ikke, de teller ordforekomster og sorterer dem statistisk. Statistikken kan deretter gjøres til gjenstand for analyse og fortolkning, men det vi fortolker da er *tall* og ikke litteratur i tradisjonell forstand. En digital metodologi vil dermed forskyve det litteraturvitenskapelige objektet fra å være bøker og forfatterskap til å bli «datasett». I artikkelen undersøker vi denne forskyvningen ved å utforske hvordan det digitale arkivet kan hjelpe oss med å studere norsk litteraturs respons på fremveksten av den kommersielle pressen i siste halvdel av 1800-tallet. Vi innleder artikkelen med å introdusere den digitale metodologien. Basert på nærlesninger fremsetter vi en hypotese om at den kommersielle pressen medførte en ny type subjektivitetsuttrykk i litteraturen. Deretter undersøker vi om det digitale arkivet avkrefter eller bekrefter denne hypotesen.

Digital humaniora

Digitale metoder i litteraturvitenskapen er en del av det som ofte kalles *digital humaniora* (DH). DH betegner ikke et eget fagfelt, men er en term som samler nye typer vitenskapelige undersøkelser og institusjonelle enheter der nye databehandlingsteknikker står sentralt (Burdick et al. 2012, 122). DH gjør seg gjeldende innenfor forskning så vel som undervisning, og det som samler feltet er langt mer metodologiske perspektiver, enn en teknologi eller et felles teoretisk utgangspunkt (Kirschenbaum 2012, 4). DH er altså ikke noe mer enn en betegnelse for bruk av digital databehandling i de humanistiske disiplinene (Berry og Fagerjord 2017). I de nordiske landene har DH lenge vært på fremmarsj, men norsk litteraturforskning har foreløpig vært avventende.² På nåværende tidspunkt er det kun det UiO-baserte forskningsprosjektet *Data-Mining the Digital Bookshelf* ved Senter for Ibsen-studier som jobber systematisk med å utvikle litteraturvitenskapelige forskningsmetoder knyttet til NBs digitaliserte boksamling.³

Digital databehandling baserer seg på kvantitative metoder som i seg selv ikke er noe nytt i litteraturvitenskapen. Tvert imot kan metodene både ses i forlengelsen av *Annales*-skolen, bokhistorie og lange tradisjoner innenfor indeksikalisering. DH står imidlertid som en naturlig fornyelse av de kvantitative metodene som oppstod med datamaskinens spede begynnelse og jesuittpresten Roberto Busas banebrytende konkordanseprotokoll over Thomas Aquinas verker, *Index Thomisticus* fra 1949. Skal vi datere faghistorien mer nøyaktig kan vi vise til tidsskriftet *Computers and Humanities* som så dagens lys i 1966. Dette tids-

1. Tilgangen er åpen så lenge man har norsk IP-adresse. Dette skiller NB.no fra mange tilsvarende digitaliserte arkiver.
2. Det finnes flere sterke DH-miljøer i Norden. Her kan trekkes frem Humlab ved Umeå universitet, Göteborgs universitets Center för digital humaniora, samt den danske DH-infrastrukturen DIGHUMLAB. Se for øvrig artikkelen «Current Efforts, Perspectives and Challenges Related to Digital Humanities in Nordic Countries» av Golub et al. (2017). Det arbeides også med DH i nordistikkmiljøene i USA. Av disse, er folkloristen Timothy Tangherlini en fremtredende utvikler av digitale metoder (Se Broadwell og Tangherlini 2018, og Tangherlini og Leonard 2013).
3. Prosjektet ledes av Ellen Rees og Frode Helland og bruker datautvinningsverktøy for å utforske rase-, kjønns-, klasse- og religionsdiskurser i det digitale arkivet.

skriftet kan sies å ha samlet fagfeltet man da kalte *Humanities computing*. Denne termen klinger ikke særlig godt i en humanists øre, og da feltet gikk fra «HC» til «DH» skyldtes det angivelig at forlaget Blackwell gikk inn for sistnevnte da de i 2004 utga antologien *A Companion to Digital Humanities*. Ifølge forlaget ville begrepet *digital humaniora* løsrive feltet fra «computing»-termen og virke mer inviterende for humanister generelt (Fitzpatrick 2012, 13).

Fra tidligere å ha vært særlig lingvistisk orientert, har den største utviklingen innenfor DH i senere tid vært innenfor litteraturforskningen. Dette har flere årsaker, men skyldes først og fremst digitaliseringens dramatiske tilgjengeliggjøring av det Margareth Cohen har kalt for det *the great unread* (Cohen 1999). Det digitaliseringen først og fremst medfører er nemlig en historisk mulighet til å kunne undersøke et bredt register av til dels nedstøvet og bortglemt materiale.

Fjernlesning vs. nærlesning?

Idéen om «det store uleste» ble popularisert av Franco Moretti omkring årtusenskiftet som på samme tid også oppfant termen «fjernlesning» («distant reading»). Ifølge Moretti utgjorde tradisjonell litteraturhistorieskrivning usikre vitenskapelige byggverk basert på lite empiri. Sannsynligvis husker og dokumenterer vi bare 1 % av det som faktisk er blitt skrevet, mens 99 % går i glemmeboken (Moretti 2000b). Morettis mål var å skrive en verdenslitteraturhistorie som representerte en større del av det skrevne materialet, og med slike ambisjoner kunne ikke forskeren forventes å ha førstehåndskjennskap til alle tekster. Tvert imot måtte man konstruere «a patchwork of other people's research, without a single direct textual reading» (Moretti 2000a, 57). Morettis begrep om fjernlesning («distant reading») innebar således en lesning av nasjonallitteraturer som baserte seg på annenhåndskunnskap, samtidig som det innebar en oppskalering av antallet tekster og forfatterskap som skulle representeres. Ideen om fjernlesning oppstod altså før massedigitaliseringen av arkivene og hadde i utgangspunktet lite med digitalisering å gjøre. Det var imidlertid først da digitaliseringen skjøt fart at ideen om en «fjernlesning» virkelig fikk gjennomslag, men nå som betegnelse for det å utforske større tekstmengder i digitale arkiver.

Særlig viktig i denne sammenhengen ble opprettelsen av *Stanford Literary Lab* som Moretti grunnla sammen med Matthew L. Jockers i 2010. Jockers har videreført Morettis kritikk av den tradisjonelle litteraturvitenskapen, og i boken *Macronalysis: Digital Methods and Literary History* (2013) mobiliserer han et solid angrep på tradisjonell nærlesning som metode. Her sammenligner Jockers studiet av litteratur med studiet av økonomi og viser til hvordan differensieringen mellom et makro- og et mikronivå endret økonomifaget etter J. Maynard Keynes' *The General Theory of Employment, Interest and Money* (1936). Jockers benytter analogien til å peke på hvordan digitaliseringen åpner et makronivå i litteraturstudiet der man kan studere litteraturens overliggende strukturer og utvikling ut fra et større materiale enn tidligere. Ifølge Jockers er nærlesning som metode uegnet for å analysere dette nye makronivået. For å gjøre det, må vi omfavne digitale metoder for innhenting og databehandling:

Like it or not, today's literary-historical scholar can no longer risk being *just* a close reader: the sheer quantity of available data makes the traditional practice of close reading untenable as an exhaustive or definitive method of evidence gathering. Something important will inevitably be missed. (Jockers 2013, 9)

Datamaskinen *fortolker* ikke, den prosesserer, og for å understreke denne forskjellen investerer Jockers i begrepet *makroanalyse* fremfor Morettis mer fortolkende *fjernlesning* (Joc-

kers 2013, 25). For Jockers representerer nemlig «analyse» også et steg mot høyere grad av objektivitet. Dette fordi nærlesning, ifølge ham, innebærer anekdotisk og spekulativ fortolkning, mens makroanalysen frembringer «comprehensive and definitive» faktakunnskap (Jockers 2013, 31).

Både Jockers og Moretti argumenterer tungt for digitale metoders høye grad av objektivitet og vitenskapelighet og deler en langt på vei positivistisk tiltro til datamaskinens analytiske overlegenhet med flere andre i DH-feltet. Ikke sjeldent kombineres denne tiltroen også med en sterk entusiasme over at digitale metoder visker ut de tradisjonelle skillene mellom natur- og åndsvitenskapene (Ganascia 2015). Mens DH-feltets forskere på den ene siden dermed har hatt en tendens til å redusere betydningen av og gjennomslagskraften til den tradisjonelt nærlesende litteraturvitenskapen, kan de på den andre siden selv anklages for å ha en vel positivistisk holdning til humanistisk forskning. Det er imidlertid nødvendig å overvinne denne (ofte svært polemiske) dikotomiseringen om vi skal nå frem til en mer *kritisk* digital litteraturvitenskap. Det motsatte av nærlesning er, som Michael Hancher har minnet om, ikke fjernlesning eller makroanalyse, men en «loose, casual, and careless reading» (Hancher 2016, 125). Formulering og testing av hypoteser går alltid begge veier og de to tilnærmingene bør derfor ikke ses i et motsetningsforhold (Olsen 2014). Tvert imot bør de kombineres for å nå frem til det den australske litteraturforskeren Katherine Bode oppsummerer som en farbar vei «beyond the view of literary history as defined by either the multiplication of data points (in Moretti's and Jocker's work) or the elaboration of unique and ultimately unknowable philological objects» (Bode 2017, 79). Bode understreker at den digitale metodologien må styres av et litteraturvitenskapelig rammeverk og at den i en slik sammenheng kan forstås som det neste logiske steget «in the cultural materialist approach» (Bode 2012, 3). Det er nettopp i et kulturmaterialistisk, og da nærmere bestemt et mediehistorisk perspektiv, at vi skal utforske den digitale metodologien i denne artikkelen.

Det sene 1800-tallet, arnestedet for vår mediekulturs subjektivitetsproduksjon?

Vår tilnærming til DH er nært knyttet til et pågående studium av hvordan norsk litteratur gjennom siste halvdel av 1800-tallet responderte på medieteknologiske endringer. Denne perioden hadde en formidabel medieteknologisk utvikling, som blant annet ga utslag i en markant avisvekst. Både i dansk og norsk sammenheng markerer årene 1848–1875 en gründerfase og gullalder for avisen (Søllinge og Thomsen 1989, 26; Dahl et al. 2010). I Norge vokste antallet aviser fra 14 i 1830 til 107 i 1880 (Dahl et al. 2010, 277). Ser vi til antallet daglige avisutgivelser i Danmark i samme periode finner vi at den femdobler seg fra 1850 til 1870-tallet (Søllinge og Thomsen 1989).

Fremveksten av den kommersielle pressen medførte en forskyvning fra en «meningspresse» til en «nyhetspresse» som i stor grad kan forstås som et resultat av utviklingen av øvrige kommunikasjonsteknologier i samme periode (Bastiansen og Dahl 2008, 213; Dahl et al. 2010, 372).⁴ Som en voksende «industri for den offentlige mening» (Bastiansen og Dahl 2008, 215) etablerte pressen et nytt kommunikasjonsregime som påvirket forholdet mellom mennesket og dets samfunnsmessige omgivelser. På den ene siden muliggjorde pressen at enkeltmenneskers handlinger og feilgrep, som tidligere hadde tilhørt privatsfæ-

4. Telegrafhistoriker Harald Rinde understreker at beretningen om skiftet fra meningspresse til nyhetspresse på 1870-tallet kan nyanseres. Flere aviser viet betydelig plass til nyhetsstoff også i tiårene før. Tross nyanseringen vektlegger imidlertid også Rinde at den elektriske telegrafens og opprettelsen av NTB bidro til å intensivere nyhetsfrekvensen i pressen (Rinde 2005, 123).

ren, nå kunne sirkuleres bredt (og hurtig) som nyhetssaker. På den andre siden forvaltet pressen individets mulighet til å kunne ytre seg i en stadig mer utvidet offentlighet. Den kommersielle pressen brakte altså med seg en ny mediesituasjon kjennetegnet ved at det moderne individet ble plassert i skjæringspunktet mellom muligheten for overvåkning/kontroll og muligheten for emansipasjon.

Ved å lese litteraturen fra slutten av 1800-tallet med et sideblikk til mediehistorien, finner vi at bøkene registrerer og problematiserer den nye mediesituasjonen. Flere sentrale tekster fra perioden beskriver mennesker som må reforhandle sin egen autonomi og selvbevissthet – sin subjektivitet – overfor pressens makt.⁵ Litterære tekster gir ofte markante uttrykk for endringer i selvbevisstheten, og flere forskere har vist hvordan institusjoner og medieteknologier bidrar til å produsere disse uttrykkene. Sentrale blant disse er for eksempel mediefilosofen Friedrich Kittler, som har fremhevet hvordan overgangen fra den tyske romantikkens idealistiske og transcendentssøkende selvforståelse til modernismens mer fragmenterte og fremmedgjorte selvbevissthet var nært knyttet til fremveksten av nye medieteknologier (fonograf og film) som gjorde det mulig for individet å erfare seg selv «fra utsiden» (Kittler 2009). Den amerikanske medieforskeren N. Katherine Hayles formulerer poenget slik: «As literary technologies change, the subjectivities they perform and inform change as well». (2008, 161)

Den teoriutviklingen som blant andre Kittler og Hayles har stått for, har vært viktig for å synliggjøre hvordan litterære subjektivitetsuttrykk er knyttet til historisk bestemte medieteknologiske substrater. Teoriutviklingen på feltet har imidlertid hatt en tendens til å overse den realistiske perioden og den medieteknologien som lå nærmest denne periodens forfattere, nemlig fremveksten av den kommersielle pressen.

Ser vi til Norge på midten av 1800-tallet er avisen ikke bare et sentralt tema i kanonlitteraturen, men de aller fleste forfatterne var i tillegg bidragsytere – på en eller annen måte – til den voksende avisfloraen (Høyer og Ihlen 1995). Og skal vi tro noen av våre mest kjente forfattere ser det ut til at etableringen av den kommersielle pressen ga grobunn til en ny type selvbevissthet basert på pressens nye muligheter for selvutøvelse og selvregulering. På den ene siden tillot den kommersielle pressen en ny type emansipasjon og selvframstillingsmulighet ved at individet enklere kunne ytre og fremstille seg i en hurtigere og langt bredere offentlighet enn tidligere, mens den på den andre siden også utgjorde en konstant trussel om å bli utlevert i den samme offentligheten.

Bjørnsons *Redaktøren* (1874/75), som gjerne omtales som «den første litterære behandlingen av redaktørrollen i Norge» (Bastiansen og Dahl 2008, 217), er et godt eksempel. *Redaktøren* kan på mange måter sies å gjøre pressen til sitt grunnleggende «problem under debatt», siden hele konflikten i stykket sentrerer rundt konsul Evjes forsøk på å unndra seg pressens makt. Ved å holde seg utenfor det offentlige ordskiftet ønsker Evje å stå utenfor byens politiske debatt. Dette tillater imidlertid ikke byens avisredaktør som skyr få midler for å få Evje til offentlig å ta den konservative fløys standpunkt:

Fruen.

De vil dog væl aldrig skrive det?

Redaktøren (ler).

Det var Johns egne ord.

5. Subjektivitet forstås i denne sammenheng som et individs utøvelse og forståelse av seg selv overfor andre. Begrepet impliserer at vår selvutøvelse og selvforståelse ikke er gitt en gang for alle, men i stadig samspill med de mennesker, institusjoner og teknologier vi omgir oss med. Subjektivitet er således noe som *produseres* i relasjonen mellom individet og dets historisk bestemte omgivelser.

Evje.

Hvad et fullt menneske siger, kommer naturligvis ikke i avisen. (til fruén). Forstår du ikke, at han spøger?

Redaktøren (kremter).

Det er allerede sat.

Doktoren.

Spøg, spøg, spøg!

Redaktøren.

Opptrinnet har givet en anonym tilstedeværende lejlighed til et meget morsomt opsæt.

Evje.

Jeg vil dog håbe –

Doktoren (til Evje).

Å, bare ærteri! Du kænner ham.

Redaktøren.

Der står: «De, som middelbart understøtter en så farlig fræmtoning, må finne sig i at blotstilles». – Jeg er enig deri.

Fruen (rejser sig).

Hvad mener De? Mener De, at Evje – ?

Redaktøren.

En smule skræk-regimente er nu ikke af vejen! (Bjørnson 1874, 31–32)

Selv om det er fullstendig nytt og uhørt for Evje, hans kone og doktoren, er det er i fullt alvor at redaktøren truer med å sitere tjeneren Johns tale fra det politiske folkemøtet i avisen. Siden John talte om «progresiv skat, om arbeids-procenter, om arbeidere som National-for-samlingens flertal» (28–29) vil dette ramme den tilsynelatende konservative konsulen om det kommer i avisen «så at alle mennesker kan læse det» (32). Da vil han fremstå som en som godtar at progressiv politikk holder hus i hans hjem, noe som naturligvis er uhørt blant borgerskapet. Når fru Evje spør om det virkelig er tilfellet at redaktøren vil løfte hennes manns privatliv inn i avisens spalter, svarer han kontant at Evje «sætter sig der selv» (34). Hva redaktøren vil eller ikke vil, er forøvrig, ifølge ham selv, et irrelevant spørsmål. Det han gjør, er kun å være «fullbyrderen af den offentlige vilje» (Bjørnson 1874, 66).

Det å kunne havne i avisens spalter var også en potensiell kilde til makt, noe Ibsen gjør til sitt hovedanliggende når Dr. Stockmann ikke får tilgang til bladet «Folkevennen» i *En folkefiende*.⁶ Den nye presseoffentligheten tematiseres også i Kristian Elsters *Farlige folk* (1881), Kiellands *Sankt Hans Fest* (1887), Garborgs *Uforsonlige* (1888) og Hamsuns *Redaktør Lynge* (1893). I sistnevnte sier redaktøren ganske enkelt: «Man har jo sine Angler ude, man dovner ikke. Man er Presse, man er Statsmagt» (Hamsun 1893, 51).

Jonas Lie innfanger dobbeltheten mer metaforisk, men med like høy presisjon i *En Malstrøm* (1884). Her utstyret nemlig Lie en ryktespredende major med avis-økenavnet «Extraposten». Når majoren beveger seg mellom whist- og kakebordet i selskap på Fossegård står det:

Og naar Rygtet i Skikkelse af Majoren saa ofte drev derigjennem fra kortbordet for at stoppe sin Pibe, havde hans stive Barteansigt Øren som Skaaler, og var hans Øjne sande Fotografiapparater. Fornøjelsen var jo ved allehaande Paafund og for højt udtalte Hemmeligheder at give «Extraposten» noget at fare med! (Lie 1884, 50–51)

6. Se (Malvik 2015).

Lies bruk av avisen som en metafor på distribuering av fortrolig informasjon kan, sammen med eksempelet fra *Redaktøren*, leses som symptomatisk for hvordan den nye mediesituasjonen ofte blir registrert negativt i perioden. Plutselig kunne det noen sa bli innfanget og formidlet til et stort publikum i en helt annen kontekst. Flere skjønnlitterære verk fra slutten av 1800-tallet gir altså inntrykk av at etableringen av den moderne presse stilte mennesket overfor et nytt og forvirrende kommunikasjonsregime der overvåking/kontroll og emansipasjon utgjorde ytterpunktene.

I vår egen tid, der disse ytterpunktene mer enn noen gang regulerer vår selvforståelse (samtidig som de regnes på av avanserte datamaskiner), oppleves det som viktig å fremskaffe kunnskap om denne formen for subjektivitetsproduksjon og dens genealogi. Spørsmålet er imidlertid om dette knippet forfattere kan sies å gi nok empiri til å generalisere. Hvordan kan vi vite at den kjente litteraturen – her representert av sju *mannlige* forfattere – er representativ for hvordan pressen bidro til å forme litterære subjektivitetsuttrykk i andre halvdel av 1800-tallet? For å svare på dette spørsmålet har vi sett et stort potensial i det å kunne konsultere NBs digitaliserte arkiv. I det følgende redegjør vi for våre fremgangsmåter og funn.

NB N-gram

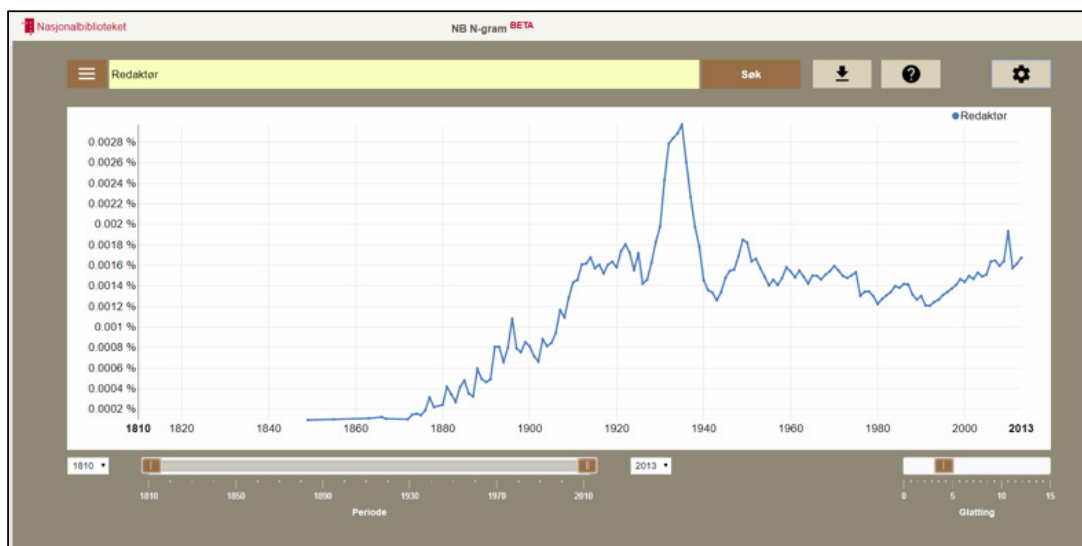
Ved å konsultere arkivet ønsker vi å undersøke 1) om det finnes eksempler utover kanonlitteraturen som behandler fremveksten av den kommersielle pressen, og 2) hvorvidt materialet støtter inntrykket fra nærlesningene av at pressens fremvekst bidro til å påvirke litteraturens subjektivitetsfremstillinger.

Det første vi gjorde for å finne svar var å ta i bruk NBs N-gram-tjeneste. Tjenesten teller ordforekomster i NBs digitaliserte arkiv av bøker og aviser og returnerer grafer hvor y-aksen angir relativfrekvensen til et valgt søkeord mens x-aksen representerer en historisk tidslinje. På bakgrunn av et ords relativfrekvens kan man dermed finne når og hvor ofte et ord forekommer i et historisk perspektiv.⁷ Prosessen er rent maskinell, og visualiseringen må fortolkes.

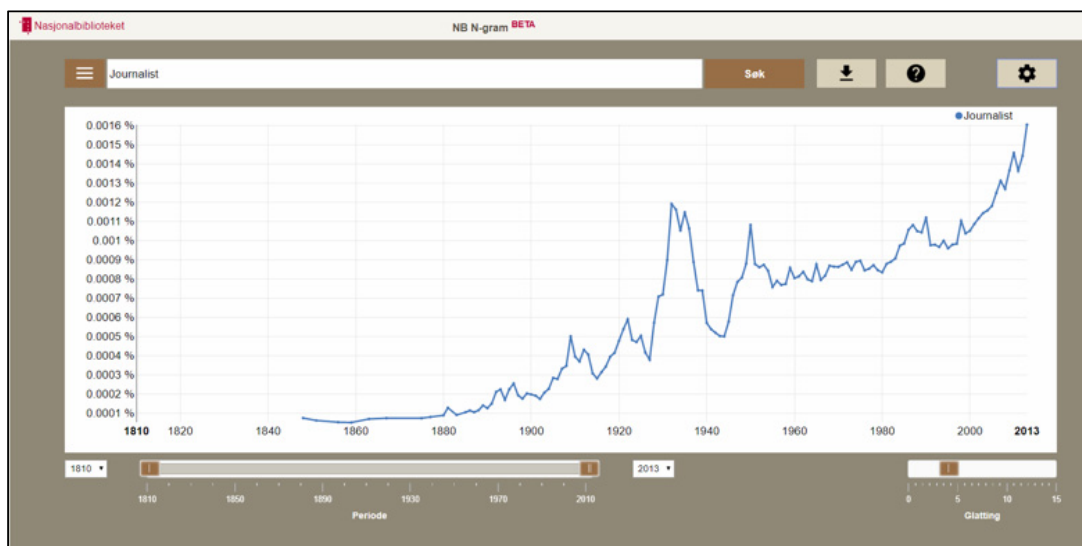
NB N-gram er et godt sted å starte for den som vil undersøke og sammenligne ords utvikling over tid. Siden vi er ute etter å avdekke noe såpass abstrakt som en ny type selvbevissthet, er det vanskelig å vite hvilke ord man skal begynne med. De litterære eksemplene ovenfor indikerer imidlertid at det nye selvforholdet var nært forbundet med pressens aktører. Vi startet derfor arbeidet med å søke etter ordene «Redaktør» og «Journalist» i bokkorpuset.

En enkel lesning av grafene (Figur 1 og 2), tilsier at ordene «Redaktør» og «Journalist» dukker opp i norske bøker rundt 1850 og blir stadig mer frekvent frem til toppunktet midt på 1930-tallet før bruken nesten halveres under annen verdenskrig. Halveringen er markant og er sannsynligvis forbundet med okkupasjonsmaktens sensur.

7. Relativfrekvensen til et ord er dets absoluttfrekvens for ett år delt på absoluttfrekvensen for alle ord i det samme året.



Figur 1 Skjermdump «Redaktør» i bokkorpuset. Hentet ut 26.06.2019.



Figur 2 Skjermdump «Journalist» i bokkorpuset. Hentet ut 26.06.2019.

Det vi er interessert i, ligger imidlertid i 1800-tallets andre halvdel, og her registrerer vi at frekvensen for begge ordene når et midlertidig høydepunkt i 1877, før den igjen stiger jevnt frem mot en toppfrekvens i 1896. I rene tall er det dette året snakk om henholdsvis 77 og 47 titler som inneholder ordene. For hele 1800-tallet (1810–1900) ligger absoluttfrekvensen på 551 «journalist»-forekomster og 773 «redaktør»-forekomster. Uten å ha beregnet hvor mange unike titler det her er snakk om – det er sannsynligvis flere bøker som forekommer i begge trefflistene, samt at N-gram ikke tar høyde for forskjellige versjoner av samme bok (som flere opplag) – gir søkene etter «Redaktør» og «Journalist» dermed et potensielt samlet korpus av 1324 bøker.

Trendlinjene i N-gram sier ingen ting om hvor mange forekomster av søkeordet som finnes i den enkelte boken. Grafene gir likevel et godt grunnlag for å hevde at det vokser frem en samtale om pressen i litteraturen fra midten av 1800-tallet og at denne samtalen for hele århundret i alle fall finner sted i 1324 bøker. Dette samsvarer med det både medie-, presse og litteraturhistorien kan fortelle om at «Bladmennene» ble en stadig mer synlig profesjonsgruppe i norsk litteratur og offentlighet (Bastiansen og Dahl 2008, 217; Dahl 1981; Dahl et

al. 2010). Søkene våre favner imidlertid langt flere titler enn dem som kommer med i historieverkene, og selv om grafenes hovedtendens er i samsvar med disse, finner vi interessante avvik. For eksempel finner vi at Bjørnson *ikke* var den første til å gi redaktøren en litterær behandling. Blant treffene, finner vi nemlig *Skitser og Fortællinger* (1866) av Hans H. Schreiber Schultz. Denne lille boken inneholder en skisse kalt «Idé» som fremstiller en redaktør under et enakter-kapittel titulert «Et Redaktionskontor».⁸

I bokens første kapittel fremtrer redaktøren på en måte som er i tråd med de senere kanoneksemlene. I en lett ironisk og humoristisk tone beskrives redaktøren både som en som «er inde i alle hemmelige Agitationer i hele Europa og Amerika» og en som vet «hvor Pengene blive av, naar det er Krise, ikke at tale om, at han ved og kjender vor egen indre Politik bedre, end baade Statsraader og Storthingsmænd» (Schulze 1866, 3–4). Schulze vektlegger også det som i vår sammenheng er det viktigste, nemlig at redaktøren er en mann som bringer det private over i det offentlig rom, og at dette får konsekvenser for hvordan mennesket forstod seg selv. Når doktoren i fortellingen skal få besøk av sin tidligere skolekamerat, redaktøren, står det nemlig:

Det var derfor ikke saa besynderligt, at den dygtige Doktorfrue feiede om og rigtig gjorde sig Umage, for at den store Mand ikke skulde faa noget at sige – eller sætte i Avisen. Doktoren smaalo og mente, at man ikke behøvede at gøre saamange Ophævelser for hans gamle Ven og Skolekammerat, som aldeles ikke var nogen extraordinær Mand i nogen Hensende. Men Fruen mente, at det før havde hændt, at man i Aviserne havde seet «Skitser fra den og den Bygd» eller «Reiseerindringer fra» o.s.v., og hun vilde aldrig kunne overleve, om der fandtes noget Slængord til hendes eller Bygdens Husstel eller Ordensands. (Vi tro, imellem os sagt, at hun havde et stille Haab om at staa *mention honorable* i den Henseende; men det var hun naturligvis for beskeden til at omtale). (Schulze 1866, 4)

Fortellingen dveler ved hvordan pressen er med på å produsere doktorfruens selvforståelse. Bare tanken på redaktørens blikk regulerer hennes adferd helt inn i det private hjemmet. Redaktøren tilskrives en maktfunksjon som gjør at doktorens kone vil fremstå best mulig. Skal vi tro fortellerens avsluttende parentes, skyldes ikke dette bare frykt for negativ omtale, men også et håp om å finne seg selv positivt fremstilt i avisen. Passasjen er således et eksemplarisk uttrykk for den nye typen subjektivitetsproduksjon generert av pressen. Schultz beskriver både muligheten til fremstilling og disiplinering av selvet, der doktorens kone regulerer seg selv på en bestemt måte relativt til redaktørens maktpotensial. Eksemplet bekrefter således inntrykket fra kanonlitteraturen.

Nøkkelordsøk og korpusbygging

N-gram-tjenesten tillater imidlertid ikke å sondre mellom skjønnlitteratur og sakprosa, og siden det først og fremst er skjønnlitteratur vi er mest interessert i, må analysen suppleres. Vårt neste skritt var dermed å bygge opp et korpus av skjønnlitterære tekster fra perioden ved hjelp av nb.no sitt «avansert søk»-grensesnitt. Ettersom vi ikke bare er opptatt av pressen, men også av 1800-tallets nye kommunikasjonsteknologier mer generelt, startet vi arbeidet med de fem søkeordene *avis**, *redakt**, *journalis**, *telegra** *fonogra** og *fotogra**.⁹ N-gram-analysen indikerte et toppår i 1896 og derfor avgrenset vi korpuset til og med dette året og søkte etter tekster utgitt mellom 1850 og 1897.

8. Schultz skriver i forordet til boken at skissene først ble publisert i tidsskriftet *Skjemt og Alvor* som han selv redigerte.

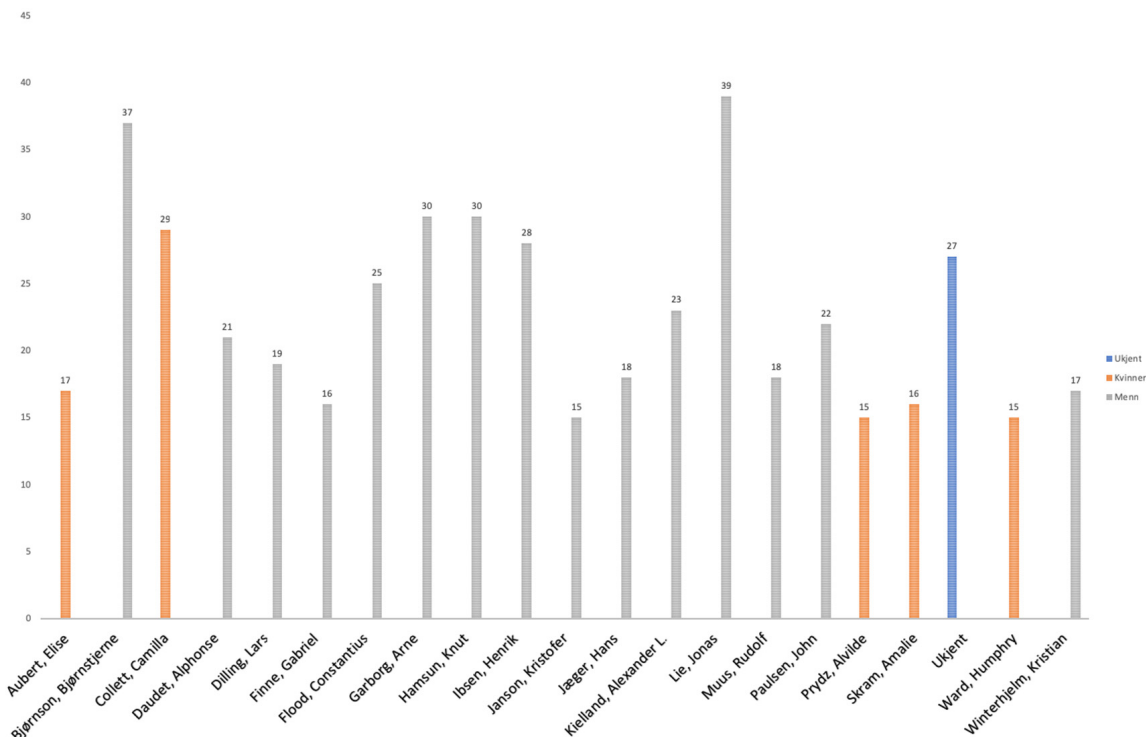
9. Asterisken bak ordet gjør at søkemotoren plukker opp flere varianter av ordene. For eksempel vil søkeordet *telegra** returnere både telegraf, telegram, telegrafistinde, telegrafen, osv.

Søkene returnerte tusenvis av titler, og som ventet var langt fra alle titlene skjønnlitterære. Dermed sorterte vi manuelt ut de titlene vi anså som litterære, og organiserte disse kronologisk etter utgivelsesår i et Excel-regneark der vi kolonneførte forfatternavn, kjønn (M/K), antall (A), tittel, utgiver, utgivelsessted og kategori (pressemedier/teknologiske medier), samt tittelens unike «adresse» (dens URN) (figur 3).¹⁰

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Søkeord	Utgitt	Forfatter	M	K	A	Tittel			Forlag	Sted	URN	Kategori
2	avis*	1851	Ibsen, Henrik	1		1	Fugl og Fuglefænger				kristiania	http://urn	Pressemedier
3	avis*	1851	Brontë, Charlotte		1	1	Shirley: fortælling, 2			Øversættelses Forlag	københavn	http://urn	Pressemedier
4	avis*	1851	Vaupell, W. L.	1		1	Soldaterliv i feltaaret 1850			Reitzel	københavn	http://urn	Pressemedier
5	avis*	1851	Rasch, Markus	1		1	Naturens vigtigste Mysterium: en Dithyrambe til Forvaringen: (trykt som Manuscript)			M. Rasch	kristiania	http://urn	Pressemedier
6	avis*	1851	Vig, Ole	1		1	Norske Bondeblomster: poetiske Forsøg			J. Hansen og O. Vig	kristiansund	http://urn	Pressemedier
7	avis*	1851	Tolderlund, Hother	1		1	Fortællinger fra Felten			Bing	københavn	http://urn	Pressemedier
8	avis*	1851	Andersen, H.C. (Hans Christian)	1		1	I Sverrig			Reitzel	københavn	http://urn	Pressemedier
9	avis*	1851	Dahl, Nils Nilssen	1		1	Guld-Thor: en Krønike i Vers			Forlaget af Udgiveren	kristiania	http://urn	Pressemedier

Figur 3

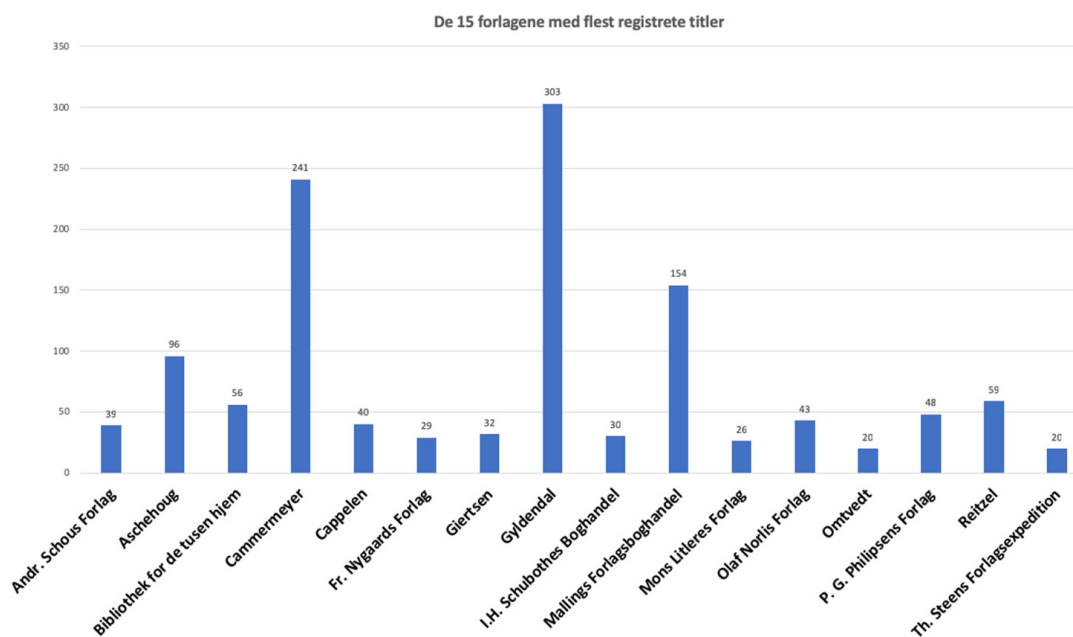
Prosedypren var fruktbar, men tidkrevende: For eksempel, returnerte søkeordet avis* 3924 treff som etter silingen ble redusert til 794 skjønnlitterære titler. Etter å ha gjennomført denne prosessen for alle nøkkelordsøkene satt vi igjen med et bokkorpus på 1864 titler. I dette korpuset gjennomførte vi enkle statistiske analyser ved hjelp av Excels pivottidogram-funksjon. Denne funksjonen viste at 350 av titlene er skrevet av kvinner, 1461 er skrevet av menn, mens forfatterens kjønn ikke lot seg registrere på de resterende 53 titlene. Nærmere undersøkelser viste at det blant de 20 mest frekvente forfatterne, altså de 20 forfatterne som har brukt ett eller flere av nøkkelordene i flest titler, ikke bare var kanonforfattere (Figur 4). Jonas Lie er forfatteren med flest registrerte titler (39), mens Bjørnson (37) følger hakk i hæl. Etter dette følger Garborg (30), Hamsun (30), Collett (29) og Ibsen (28) tett på, mens Kielland, Hans Jæger (18) og Skram ligger godt bak med henholdsvis 23 og 16 titler.



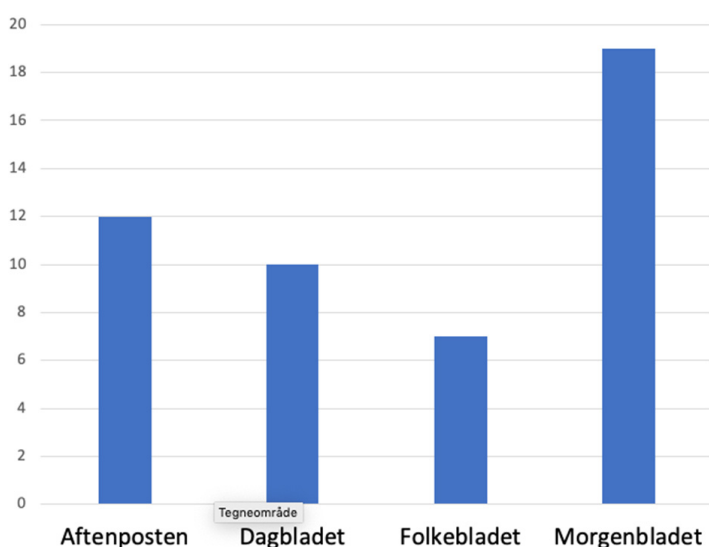
Figur 4

10. Innsamlingen, sorteringen og organiseringen av korpuset ble utført av Boasson og Malvik sammen med vit.assene Vilde Nybakken og Marie Skare Irgens.

Kanonforfatterne er altså godt representert med til sammen 232 registreringer. Utover dette peker listen imidlertid ikke bare til Constantius Flood (25), Lars Dilling (19), Elise Aubert (17), Kristian Winterhjelm (17), Rudolf Muus (18), Gabriel Finne (16), Kristofer Janson (15) og Alvhild Prydz (15). Den synliggjør også at det ble utgitt en betydelig del oversatt litteratur. For eksempel er den franske realisten Alphonse Daudet og den viktorianske pedagogen og romanforfatteren Mary Augusta Ward (under pseudonymet Humphry Ward) registrert med henholdsvis 21 og 15 titler. Tallene indikerer også en markant bredde i hvordan litteratur ble distribuert (Figur 5). Eksempelvis markerer avisene seg som trykkere av egne føljetonger med for eksempel *Morgenbadet*, *Aftenposten*, *Dagbladet* og *Folkebladet* representert med henholdsvis 19, 12, 10 og 7 titler hver (Figur 6).



Figur 5



Figur 6

Det ser altså ut til å være langt flere tekster fra perioden enn bare kanonverker som omhandler aviser, redaktører og nye medieteknologier, samt at disse tekstene like gjerne ble distribuert lokalt og gjennom pressen selv.

Problemene står imidlertid i kø når man vil hente ut pålitelig statistikk fra dette datasettet. For eksempel er det, her som ellers i forbindelse med det digitale arkivet, store problemer med OCR-feil. OCR står for *optical character recognition* og er den programvaren som sørger for at bokstavene på originalsiden blir representert korrekt i den skannede kopien. Det er langt fra sjelden at de skannede representasjonene er gale, og da særlig i forbindelse med bokstavene Æ, Ø og Å. Dermed er det også relativt sett stor sannsynlighet for at mange ord som skulle blitt fanget inn ikke blir det, fordi de ikke er representert (stavet) korrekt i det digitale arkivet.

En annen utfordring er at nb.no rommer flere utgaver av bøkene og dermed vil samme tittel telles flere ganger. Dette løste vi ved å manuelt sile ut slike forekomster, men siden de fleste titlene som inneholder ordene «redakt*» eller «journalis*» også inneholder «avis*» og/eller «telegra*» hadde vi fortsatt mange titler som var registrert flere ganger i vårt samlede korpus.¹¹ Selv om vi har gått nøye gjennom datasettet, er det altså knyttet store feilmarginer til korpuset som gjør det vanskelig å tolke frem pålitelige resultater. Likevel tror vi at det er mulig å lese ut et overordnet bilde som viser at litteraturens registrering av periodens medieendringer også må søkes ut over kanon, og da både i oversatt og lokalt distribuert litteratur.

Gevinsten av vår korpusbygging er i første omgang ikke knyttet til pålitelig statistikk, men til selve listen over de 1864 titlene som kan relateres til fremveksten av en ny mediekultur. Denne listen danner nemlig et godt utgangspunkt for å gjøre et bredere nedslag i hvordan både kjente og ukjente litterater knyttet nye medieteknologier til nye former for selvbevissthet på 1800-tallet. Ved hjelp av dette korpuset blir vi for eksempel minnet om at Camilla Collett sammenlignet en teksts røde tråd med den telegrafiske atlanterhavskabelen i *I de lange nætter* (1863), og vi lærer om hvordan oppfinnelsen av telegrafen tillot sogneprest Sigismund Christian Dick (1815–1882) en ny måte å eksternalisere den indre protestantiske kommunikasjonen med Gud.¹²

Dick ble sogneprest på Aukra i 1857, og i salmen «Telegrafen» fra 1861 skriver han:

Hurtig Telegrafen bringer / Bud til fjerne Steder hen / Fuglen ei paa lette Vinger / naar saa hastig Him-
melen. / Snarere dog Bønner stige / op til Lysets fjerne Rige / til Guds aabne Faderbryst. // Er fra dit til Jesu
Hjerte / fæstet Troens stærke Traad, / o, da veed du i din Smerte / dog det allerbedste raad. / Bønnens Tele-
graf din Klage / bringer frem og Svar tilbage, / og af Herren Trøst du faar. (Dick 1861, 25–26)

Salmens første strofe setter opp en kontrast mellom en horisontal bevegelse, telegrafens evne til lynraskt å mediere informasjon på tvers av store avstander, og en vertikal bevegelse, bønnens umiddelbare oppkobling til Guds faderbryst. I andre strofe kobles kommunikasjons teknologi og ånd sammen av metaforene «Troens stærke Traad» og «Bønnens Tele-

11. Dette forklarer hvorfor for eksempel Lars Dilling kommer med på listen over de 20 mest frekvente forfatterne i perioden. Dilling skrev en del mindre bøker, blant annet *Hverdagsmennesker: Skitser* som kom i tre ulike bind og *Gjennom Lorgnetten: Skitser* som også fyller tre bind. Når alle disse 6 bindene inneholder flere av nøkkelordene har de til sammen blitt registrert 17 ganger. I tillegg har *Fra Tyskland og Belgien: Skitser* blitt registrert to ganger. Dermed har de syv reelle titlene blitt registrert 19 ganger, noe som igjen gir Dilling en høy total representasjon i korpuset. Vi finner et tilnærmet likt forholdstall når det gjelder Joans Lie der 17 unike titler til sammen har blitt registrert 39 ganger.

12. Collett skriver i *I de lange nætter* at den røde tråden som hun mente gikk gjennom hennes ungdoms brev og dagboksnotater i 1863 fremstod like avrevet for henne som «den undersøiske Telegrafstreg, der brast netop som man drømte den stolteste Drøm om at forene tvende Halvkugler» (Collett 1863, 2).

graf», i et poetisk bilde av det protestantiske subjektets kommunikasjon med Gud. Dette eksempelet tydeliggjør hvordan telegrafens tilbød en ny måte å konseptualisere kommunikasjonen med Gud på, samtidig som det blir klart at telegrafens representerte en form for felles kunnskap også i den geografiske periferien Aukra bare seks år etter at utbyggingen av telegrafens startet i Norge.

Både Collett og Dick bruker den elektriske telegrafens som bilde på subjektets evne til å formidle sin tankekraft. Eksemplene gir således grunn til å bygge ut undersøkelsen av relasjonen mellom de nye kommunikasjonsteknologiene og periodens skiftende forståelse av selvet. Vi støter imidlertid også her på en rekke problemer. Et nøkkelordsøk returnerer nemlig treff uavhengig av hvilken sammenheng nøkkelordet opptrer i, og hvilken funksjon det har i setningen. Et ordsøk sier dermed ingenting om *bruken* av ordet, eller om meningen til setningen det er en del av. Dette må man fortsatt manuelt gå inn og undersøke. Når listen enkeltverk begynner å romme et par tusen titler, blir jobben relativt omfattende og tidkrevende, og det søkene medfører i praksis, er dermed i all hovedsak å løfte frem «det store uleste» uten å hjelpe oss i analysen. I tillegg er det ingen garantier for at søkene returnerer de mest interessante eksemplene vi er ute etter. Slik det metaforiske eksempelet fra Lies *En Malstrøm* synliggjør, kan en tekst kritisere avismediet uten at ordet avis nevnes. Denne utfordringen med ordsøk fremheves også av Timothy Tangherlini og Peter Leonard, som poengterer at selv ikke en ekspert på et felt vil kunne forutsi alle de ordene som potensielt kan brukes for å beskrive temaet vi er ute etter å undersøke (Tangherlini og Leonard 2013, 728). Hverken nøkkelordsøk eller N-gram kan altså fortelle oss noe om hvilken kontekst ordene opptrer i og slik heller ikke noe om hvordan de ble brukt.

Kollokasjonsanalyse i Jupyter Notebook

Nasjonalbiblioteket har gjennom de siste årene utviklet en rekke verktøy som kan brukes til å analysere den digitaliserte boksamlingen deres. Forskningsbibliotekar Lars Johnsen har for eksempel utviklet et programmeringsgrensesnitt (API) som muliggjør kollokasjonsanalyser i det digitale arkivet ved hjelp av verktøyet Jupyter Notebook.¹³ En kollokasjon er en språkvitenskapelig term som angir en «forbindelse av to eller flere ord som vanligvis opptrer sammen, for eksempel *gjøre fremskritt* og *felle en dom* (NAOB «kollokasjon» 2018). Analyseverktøyet registrerer og teller opp den semantiske konteksten til et ord man er opptatt av. For eksempel kan vi telle opp hvilke ord «Avis», «Redaktør» og «Journalist» vanligvis opptrer sammen med i gitte perioder. En slik opptelling vil altså kunne gi indikasjoner på hvordan søkeordene ble brukt, og slik også hvordan de nye medieteknologiene ble beskrevet og oppfattet i litteraturen.

Jupyter Notebook er en *open source*-webapplikasjon hvor brukere kan lage og dele dokumenter som inneholder programmeringskode, likninger, visualiseringer og vanlig tekst. Johnsen har utviklet dette grensesnittet opp mot både bok- og avisarkivet og gjort det mulig å hente ut for eksempel konkordanselister, plott vekstdiagram, sammenligne ulike korpus, og undersøke OCR-feil, samt bestille den typen kollokasjonsanalyser som vi her er på jakt etter.¹⁴ I tillegg til å returnere en liste over kollokasjoner, generer programmet et referansekorpus slik at kollokasjonene kan rangeres i henhold til en relativfrekvensverdi. Mens N-gram-analysene og nøkkelordsøkene våre indikerer at det tar til en «samtale» om medieteknologier omkring 1850-tallet, kan kollokasjonsanalysen altså samle mer komplekse

13. Johnsen har fungert som konsulent for arbeidet med Jupyter Notebook som ligger til grunn for denne artikkelen.

14. Av litteraturvitenskapelige miljøer i Norge, er det foreløpig forskningsprosjektet *Data-Mining the Digital Bookshelf* ved Senter for Ibsen-studier som jobber tettest med NB om å bruke og utvikle disse verktøyene.

semantiske strukturer enn bare enkeltord. På denne måten kan vi få en fornemmelse av hvordan ordene ble brukt, for eksempel positivt eller negativt, og slik har vi også et mer raffinert verktøy for å avdekke hvordan fremveksten av subjektivitetsuttrykk i kjølvanet av pressens utvikling er registrert i det digitale arkivet.

«Redaktørens» og «Journalistens»

Nærlesninger av kanonlitteraturen og mer ukjente eksempler, som Schultze, viser at det i beskrivelser av redaktører og journalister ligger informasjon om hvordan «vanlige folk» regulerte sin egen adferd overfor bladfykene. Dermed er det nærliggende å tro at en større undersøkelse av hvordan redaktører og journalister ble beskrevet kan føre oss nærmere et svar på om og hvordan deres subjektivitetsproduserende funksjon er registrert i det digitale arkivet. På bakgrunn av dette har vi gått ut fra at en kollokasjonsanalyse kan gi oss et kvantitativt anslag på hvilke ord forfattere brukte til å beskrive pressefolkene. Slik ble det viktig for oss å finne ord som var semantisk attribuert til redaktører og journalister, og vi falt derfor ned på at ord som forekommer sammen med *genitivs*formene, «redaktørens og journalistens», burde være av særlig interesse. S-genitiv uttrykker tilhørighetsforhold, og sjansen er derfor større for at de ordene som kollokasjonsanalysen plukker opp faktisk er semantisk attribuert til, og altså «handler om», redaktører og journalister. Analysen skulle på denne måten kunne gi en indikasjon på historiske holdninger til de to profesjonene.

I tillegg til å få en indikasjon på hvilke ord som opptrer sammen med «Redaktørens» og «Journalistens» i perioden, valgte vi å dele opp søket i tidsintervallene 1850–1870, 1870–1890 og 1890–1910, for kunne spore en historisk utvikling.¹⁵ Vi begrenset konteksten til å inkludere ett ord før nøkkelordet og to ord etter. Dersom korpuset for eksempel inneholdt setningen «den smarte redaktørens voldsomme makt», ville verktøyet dermed returnere adjektivet «smarte», som står attributivt til «redaktørens», adjektivet «voldsomme», som står attributivt til «makt», og «makt», som er kjernen i substantivfrasen.

Nedenfor lister vi opp fire tabeller (Figur 7) som viser de 20 mest frekvente ordene rundt nøkkelordet «Redaktørens». Tabell 1 og 2 (fra venstre) er begge avgrenset til perioden 1850–1870. Dette kommer av at ordet redaktør ofte forekommer som «Redacteur» i denne perioden. Tabell 3 og 4 angir kollokasjoner fra henholdsvis 1870–1890 og 1890–1910. I første omgang ser det ut til at tabellene viser hvordan den gradvise profesjonaliseringen av redaktøryrket også ga seg utslag i det språket man brukte for å beskrive dem. Ordene «Avis» og «egen» i første kolonne vitner om en tid da redaktører gjerne tok seg av alle avisens oppgaver på egenhånd. Beveger vi oss inn i tabell 3, 1870–1890, finner vi at flere ord angir de fasilitetene som står til redaktørens disposisjon, slik som «Værelser», «Privatsekretær», «Pult» og «kontor», ord som tilsynelatende løfter frem veksten av moderne redaksjonslokaler med flere ansatte. Den siste kolonnen ser ut til å bekrefte dette forholdet ettersom ordet «Kontor» nå er det mest frekvente.

15. Ettersom vi ønsket å spore utvikling over tid, var det viktig at tidsintervallene kunne sammenlignes med hverandre. Vi utvidet derfor tidsrommet fra de tidlige ordsøkene slik at vi nå fikk tre 20-årsintervaller.

	Redaktørens		Redacteurens		Redaktørens		Redaktørens
Avis	1.605483	altid	1.525530	til	1.429575	Kontor	1.385186
egen	1.605483	mod	1.525530	Knaphul	1.416442	til	1.362793
Hustru	1.605483	fra	1.525530	private	1.416442	over	1.321588
vinde	1.605483	ikke	1.525530	Pult	1.321588	paa	1.321588
Bifald	1.605483	Talent	1.525530	Kontor	1.321588	fra	1.292424
havde	1.605483	Side	1.525530	tilbage	1.321588	Haand	1.292424
i	1.100736	Navn	1.525530	Læte	1.321588	da	1.255773
at	1.018944	Mening	1.525530	Person	1.321588	ikke	1.255773
og	0.760526	Hr.	1.525530	Privatkontor	1.321588	private	1.255773
jeg	0.000000	Forpligtelse	1.525530	Privatsekretær	1.321588	Privatkontor	1.255773
in	0.000000	er	1.428895	alle	1.321588	Person	1.255773
hun	0.000000	og	0.658169	Udtalelse	1.321588	eget	1.255773
!	0.000000	.	0.552916	Vaabensvend	1.321588	under	1.255773
han	0.000000	,	0.222300	Vcerelser	1.321588	Stemme	1.205874
et	0.000000	jeg	0.000000	Vidende	1.321588	Leder	1.205874
er	0.000000	det	0.000000	Værelse	1.321588	Han	1.205874
en	0.000000	die	0.000000	Yærelse	1.321588	Handlemaade	1.205874
du	0.000000	»	0.000000	temmelig	1.321588	Hr.	1.205874
f	0.000000	hun	0.000000	Sinn	1.321588	Hukommelse	1.205874
la	0.000000	i	0.000000	Hr.	1.321588	I	1.205874

Figur 7 Rangering av hvilke ord som oftest forekommer sammen med ordet «redaktørens». De to første tabellene fra venstre viser resultater fra perioden 1850–1870. De to siste viser henholdsvis periodene 1870–1890 og 1890–1910.

Men her finnes også ord som peker i retning av at avisfolk ble beskrevet som maktpersoner i litteraturen. Ordene «Bifald», «Navn», «Mening» og «Forpligtelse» fra den første perioden, «Vaabensvend», «Udtalelse» og «Person» fra den andre perioden, og «Haand», «Leder» og «Stemme» fra den siste, støtter fremstillingen av redaktøren slik den kommer til uttrykk i kanonverkene. Ser vi på samme analyse gjort med «Journalistens» (Figur 8) bekreftes inntrykket av at 1800-tallets litteratur tilskrev pressens menn en betydelig maktfaktor. Her fester vi oss ved ordet «Domstol», som forekommer både i første og siste periode, men vi legger også merke til at ord som «Kritik», «kaade», «slibrige», «Magt», «despotiske» og «kolde» er særlig hyppige. De fleste av disse forekommer i perioden 1870–1890 og indikerer således at bokdiskursens «samtale» var kritisk til den fremvoksende pressens nye makt. Det er også nærliggende å tro, særlig grunnet kollokasjonene «Domstol», «Kritik», «Magt» og «despotiske», at makten er skildret som en faktor som medfører enkeltmenneskets regulering og korrigerende av egen atferd og selvbevissthet, altså som subjektivitetsproduserende maktformer slik vi også finner dem i kanonverkene eller hos Schultze der doktorens kone regulerte sin egen atferd i både håp og frykt for å få en positiv «dom» i avisen.

Tabellene ser ut til å gi grunnlag for å underbygge påstanden om at fremveksten av den kommersielle pressen medførte nye subjektivitetsuttrykk i litteraturen. Et problem med disse resultatene er imidlertid at de baserer seg på *relativ*frekvenser og at relativfrekvenser skjuler absolutte tall. Undersøker vi de absolutte tallene for analysen finner vi at grunnlaget for tabellene er to/tre bøker (Redaktør/Journalist) for perioden 1850–1870, 19/12 bøker fra 1870–1890, og at det først nærmer seg et større antall når vi kommer frem til 1890–1910, der funnene baserer seg på 67/40 bøker. Når grunnlaget for den første perioden viser seg å

være noen få bøker, som vi attpåtil kjente til fra før, er det heller ikke rart at resultatene er i overensstemmelse med det vi allerede visste. Dermed er datagrunnlaget for tynt til å kunne bekrefte noe som helst før vi kommer til den siste perioden der det med en viss rett kan hevdes at vi når ut over hva vi ville gjort ved hjelp av den tradisjonelle nærlesningen.

	Journalistens		Journalistens		Journalistens
tværtimod	1.605483	af	1.528675	:	1.418009
Domstol	1.605483	Kald	1.511060	jus	1.403615
hvilket	1.605483	altid	1.511060	Pen	1.403615
pna	1.605483	sikkert	1.511060	med	1.374523
hurtige	1.605483	Smule	1.409868	vitæ	1.363811
stiller	1.605483	Broehurer	1.409868	Domstol	1.363811
Fortale	1.605483	Kritik	1.409868	skiller	1.363811
og	0.880861	kaade	1.409868	vi	1.363811
.	0.484178	Tilsvar	1.409868	ogsaa	1.363811
for	0.000000	slibrige	1.409868	kolde	1.363811
f	0.000000	besad	1.409868	Liv	1.309619
hun	0.000000	iuZ	1.409868	da	1.309619
et	0.000000	Magt	1.409868	hørt	1.309619
er	0.000000	ikke	1.409868	hvilken	1.309619
en	0.000000	Virksomhed,	1.409868	Vine	1.221918
det	0.000000	vitso	1.409868	sædvanlige	1.221918
i	0.000000	despotiske	1.409868	liv	1.221918
!	0.000000	Politikerens	1.409868	Redegjørelse	1.221918
of	0.000000	Evne	1.409868	Tale	1.221918
jeg	0.000000	til	1.405175	Tilsvar	1.221918

Figur 8 Rangering av hvilke ord som oftest forekommer sammen med ordet «journalistens». Fra venstre til høyre viser tabellene periodene 1850–1870, 1870–1890 og 1890–1910.

Det funnene så langt underbygger, er altså ikke nødvendigvis at hypotesen er riktig, men at metoden i alle fall ikke er gal: Basert på det vi kjente, bekreftet metoden det vi allerede visste. Vår prøving og feiling med kollokasjonsanalyser stiller oss således overfor en metodologisk utfordring: Ved å telle lavfrekvente ord ser det ut til at vi i stor grad vises tilbake til de tekstene som i utgangspunktet fremprovoserte hypotesen, altså tilbake til kanon.

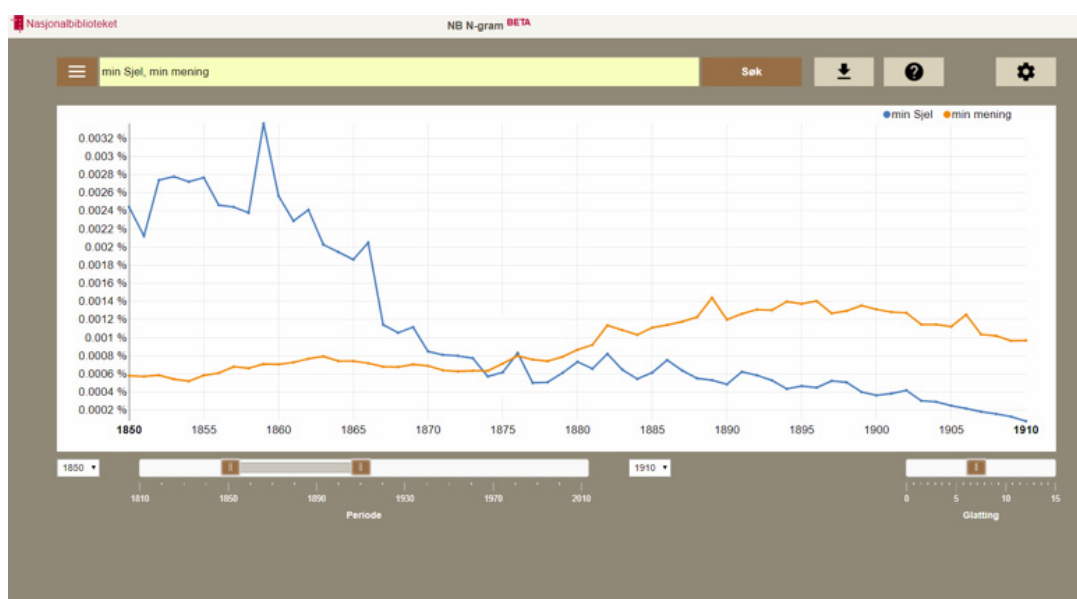
«min Sjæl» vs. «min Mening»

For virkelig å konsultere «det store uleste» må vi gå ut over lavfrekvente nøkkelord som «Redaktørens» og «Journalistens» og heller fokusere på høyfrekvente ord. Neste steg i vår analyse ble derfor å gjennomføre en serie kollokasjonsanalyser av ordet «min». «Min» er en determinativ possessiv og brukes som «regulær eieform til jeg, for eksempel 'min sykkel', 'kjæresten min', 'min tanke', osv.» (NAOB «min» 2019). «Min» artikulere et eiendomsforhold mellom et jeg og dets fysiske og begreplige omgivelser. Vår idé ved å gjøre «min» til et nøkkelord i kollokasjonsanalysen, var at ord som står sammen med ordet «min» ofte vil være ord som er possessivt attribuert til selvet, og at denne determinativens kollokasjoner

dermed ville kunne gi en indikator på hvordan forfattere har fremstilt menneskets selvforhold opp gjennom 1800-tallet. Tanken var at dersom vi kunne registrere historiske endringer i hvilke ord som vanligvis sto sammen med «min», ville vi ha et statistisk mål på historiske svingninger i bokkorpusets subjektivitetsfremstilling. Hvis vi i tillegg kunne knytte slike svingninger til mediehistoriske endringer, ville vi ha et langt bedre grunnlag for å kunne generalisere våre antakelser om at periodens medieendringer avstedkom endrede subjektivitetsuttrykk i litteraturen. Denne gang brukte vi absoluttfrekvenser. Vi søkte i alle tilgjengelige bøker fra de samme tre periodene og reduserte konteksten til ett ord foran og ett ord etter nøkkelordet «min». I 3051 bøker fra 1850–1870 fant vi 59.627 unike kollokasjoner, i 5456 bøker fra 1870–1890 fant vi 66.067 unike kollokasjoner, mens 10.459 bøker fra 1890–1910 rommet 99.696 unike kollokasjoner.

Tallmaterialet er et ganske annet enn i forbindelse med redaktørene og journalistene. Når vi så begynte å sammenligne listene over kollokasjoner, oppdaget vi noe interessant: kollokasjonene «min Sjel» og «min Mening» så ut til å stå i et omvendt proporsjonalt forhold til hverandre. I perioden 1850–1870 finner vi 2844 forekomster av ordet «Sjel», mens ordet «Mening» forekommer 1977 ganger. I neste 20-årsperiode er antallet «Sjel»-kollokasjoner redusert til 1466, mens «Mening» har steget til 2521. I siste periode, 1890–1910, har «Mening» kommet opp i 4940, mens «Sjel» ikke er å finne på listen over de 60 mest frekvente kollokasjonene.

Et kontrolløk etter frasene «min sjel» og «min mening» i NB N-gram bekreftet i første omgang denne omvendte proporsjonaliteten, skjønt grafene virket nærmest å være for gode til å være sanne. Det grafene viste var nemlig at frekvensen for «min Sjel» sank dramatisk fra den romantiske perioden på midten av århundret og frem mot århundreskiftet, mens frasen «min Mening» steg jevnt og trutt i takt med utviklingen av den kommersielle pressen. Ikke minst viste trendlinjene, helt i tråd med presse- og litteraturhistorien, et taktskifte i begynnelsen av 1870-årene der «min Mening» i året 1874 – symbolsk nok i samme år som Bjørnsøns *Redaktøren* – for første gang ble mer frekvent enn «min Sjel» (Figur 9).



Figur 9 Skjermdump «min sjel» og «min mening» i bokkorpuset. Hentet ut 26.06.2019.

1870-tallet er det tiåret da infrastrukturer for offentlig meningsutveksling for alvor gjør seg gjeldende i Norge. NTB ble opprettet i 1868 og i 1870 nådde telegrafutbyggingen Hammer-

fest. Samme år besørget marineminister Ole Jacob Brochs reformer i postvesenet som sikret daglig postgang på hovedrutene i det sørlige Norge (Seip 1971, 451–452). Denne samferdselsutviklingen innebar at pressen i større grad enn tidligere kunne formidle oppdatert nyhetsstoff (Bastiansen 2001, 40). Sett med våre øyne så dermed dreiningen i bokkorpuset fra «min Sjel»-kollokasjoner til «min Mening»-kollokasjoner ut til å sammenfalle med fremveksten av en moderne medieoffentlighet i Norge, noe som igjen bekreftet antagelsen om at medieutviklingen ga seg utslag i litteraturens subjektivitetsframstillinger.

Men også her måtte vi ta ett skritt tilbake. Nærmere studier viste nemlig en rekke OCR-feil. Både «Sjel» og «Mening» finnes nemlig stavet i flere varianter i bokkorpuset.¹⁶ Dette fikk størst utslag i antallet «Sjel»-kollokasjoner, som nå også måtte inkludere «Sjæl», «Sjcel», «Sjcr1», «sjel» og «sjæl». «Mening» ble ikke skrevet med æ og er dermed mindre utsatt for OCR-feil, men forekommer likevel til dels med to e-er, «meening», i tillegg til de to variantene med stor og liten forbokstav. Dette betydde altså at tallene våre var gale, og at vi måtte kjøre søket igjen med høyde for alle mulige varianter av ordene.

For å løse dette problemet lagde vi en Sjel-variabel som samlet alle stavevarianter av ordet «Sjel» og en mening-variabel som samlet alle stavevarianter av «mening», og så gjorde vi kollokasjonsanalyser hvor ordene «sjel» og «mening» – altså ikke «min» – var nøkkelord. Vi søkte bare etter det ordet som står rett før sjel/mening fordi vi nå var på jakt etter frasene «min sjel» og «min mening». Dermed kunne vi justere funnet. I absolutte tall fant vi nå at frasen «min sjel» forekom oftere enn «min mening» gjennom hele perioden. Grafen som viste et litteraturhistorisk skifte fra «min sjel» til «min mening» akkurat i året 1874 viste seg således for god til å være sann. Dette betyr imidlertid ikke at kollokasjonsanalysen motbeviser N-gram-analysen. Den nyanserer den. Å lese kvantitativt, er å lese tall, og før vi forkastet analysen måtte vi se nærmere på tallene.

Rubrikken nedenfor viser tre regnestykker hvor absoluttfrekvensene for «min sjel» deles på absoluttfrekvensene for «min mening» i hver 20-årsperiode. Hvis vi starter med å se på absoluttfrekvensene, altså det totale antallet forekomster innenfor den definerte tidsintervallen, ser vi at begge frasene tiltar i frekvens frem mot århundreskiftet. Denne økningen indikerer sannsynligvis at det er flere digitaliserte bøker i arkivet fra 80-tallet enn fra 60-tallet osv.

Periode 1850–1870: «min sjel» 8946 / «min mening» 2246 = 3,9

Periode 1870–1890: «min sjel» 9010 / «min mening» 2737 = 3,29

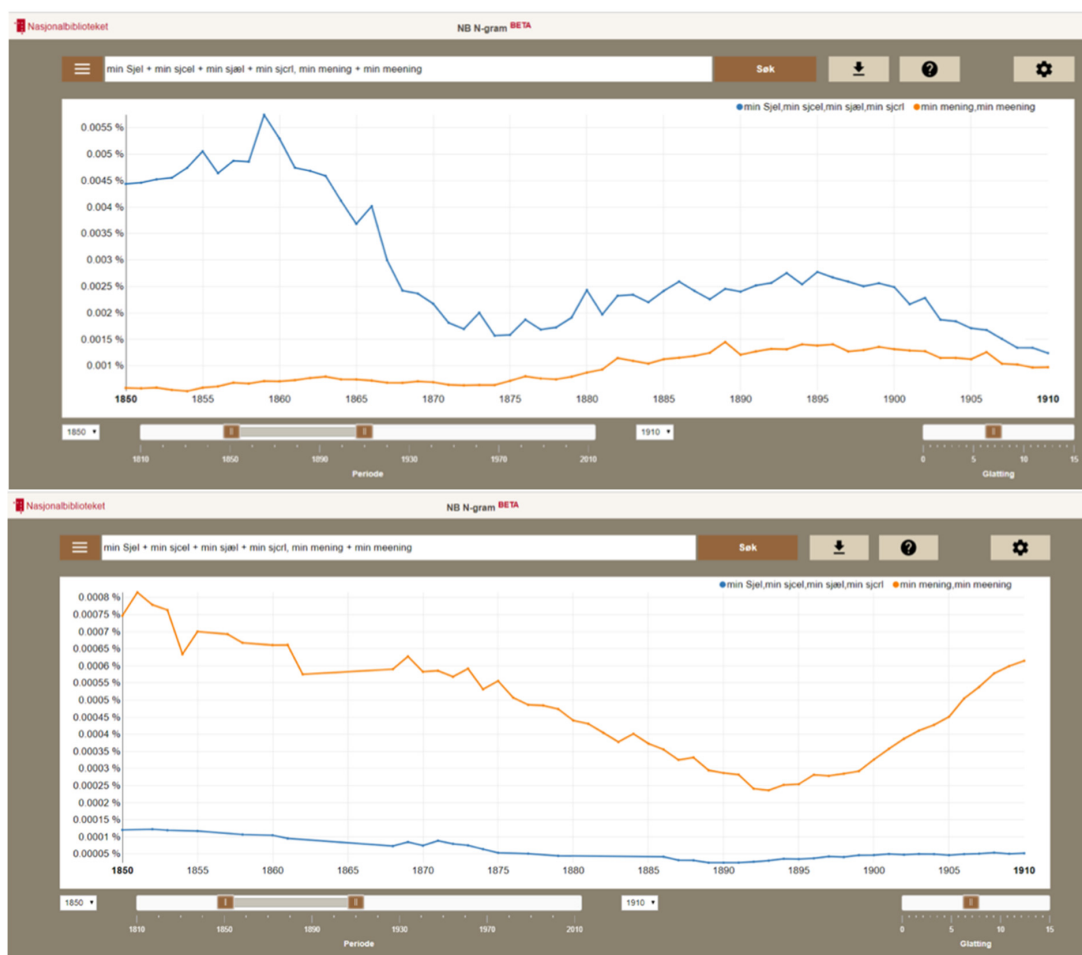
Periode 1890–1910: «min sjel» 14689 / «min mening» 8259 = 1,77

Men hvordan utvikler de to frasene seg i forhold til hverandre? Dette finner vi ut ved å dele antallet «min sjel»-kollokasjoner på antallet «min mening»-kollokasjoner. Svaret på regnestykket gir en proporsjonal verdi for hver periode. Interessant nok, ser vi at denne verdien minker fra periode til periode. Dette betyr at relativt til forekomsten av «min sjel»-kollokasjoner, så er «min mening»-kollokasjonene på fremmarsj gjennom 1800-tallets siste halvdel. Tallene fremhever en klar trend der «min sjel» og «min mening» utvikler seg omvendt proporsjonalt i forhold til hverandre gjennom 1800-tallets andre halvdel. Med andre ord, bekrefter den kvantitative undersøkelsen våre antagelser fra nærlesningene, at litterære subjektivitetsuttrykk – gitt at frasene «min sjel» og «min mening» er litterære subjektivitets-

16. For å finne ut av hvilke varianter av feilstavelser som finnes av et ord, brukte vi en såkalt «wild card-algoritme» som Lars Johnsen utviklet i forbindelse med Heidi Karlsens pågående Ph.d.-prosjekt. Karlsen forsker på kvinne-diskurser i bokkorpuset fra 1800-tallet.

uttrykk – endrer seg parallelt med fremveksten av den kommersielle pressen, om enn ikke på en like spektakulær måte som N-gram-analysen så ut til å implisere.

Disse funnene blir enda mer interessante om vi gjør samme analyse i det digitaliserte aviskorpuset. I avisene er nemlig forekomsten av «min mening» høyere enn forekomsten av «min sjel» gjennom hele perioden, og korrigererte grafer i N-gram for henholdsvis bok- og aviskorpuset ser slik ut (Figur 10):



Figur 10 Den øverste grafen viser forholdet mellom «min sjel»/«min mening» i bokkorpuset, mens den nederste viser resultatet fra aviskorpuset. I bokkorpuset har «min sjel» (blå linje) en nedadgående kurve i forhold til «min mening» (oransje linje), mens i aviskorpuset er «min mening» gjennomgående mer frekvent enn «min sjel». Skjermdumpene er hentet ut 26.06.2019.

Disse grafene baserer seg på 32.645 forekomster av «min sjel» og 13.242 forekomster av «min mening» i det samlede bokkorpuset bestående av 21.332 bøker. Etter en del prøving og feiling har vi dermed kommet frem til en kvantitativ analyse som viser at «min sjel» taper terreng til «min mening» frem mot århundreskiftet, i tillegg til at «min mening» er gjennomgående mer frekvent enn «min sjel» i aviskorpuset.

Hvordan fortolker vi disse funnene? Utviklingen av den omvendte proporsjonaliteten mellom frasene «min sjel» og «min mening», som vi finner i bokkorpuset, kan se ut til å ta etter et diskursivt forhold som allerede finnes i avisene. Kanskje betyr det at vi har å gjøre med en type diskursiv migrasjon fra presse til litteratur? I så fall, hvilken rolle spilte egentlig pressen i den «avsjelingen» i litterære subjektivitetsuttrykk som kollokasjonsanalysen antyder? Finnes det andre kollokasjonsutviklinger som peker i samme retning? Grafene gir neppe grunnlag til å trekke utvetydige historiske konklusjoner. De åpner likevel for en del

spørsmål som kan ta studiet av litteratur og dens medieteknologiske omgivelser videre.

Som subjektivitetsuttrykk betraktet, er «min sjel» interioriserende mens «min mening» er eksterioriserende. I en idé- og litteraturhistorisk fortolkningsramme er det fristende å se det første som uttrykk for romantikkens transcendentssøkende subjektivitet og det andre som uttrykk for den realistiske periodens problemorienterte og offentlig meningsbrytende subjektivitet. Funnene forteller oss i så fall, og ikke overraskende, at det man ofte anser som romantiske subjektivitetsuttrykk gjorde seg gjeldende lenge etter 1850, mens den meningsbrytende tendensen er et resultat av en langt mer kontinuitetsbærende prosess enn det som ofte kommer til uttrykk i litteraturhistoriens fortelling om Brandes og doktrinen om å «sette problemer under debatt».

Litteraturhistoriske periodebetegnelser kan representere grove forenklinger av komplekse utviklingsprosesser, og det er her nærliggende også å knytte funnene våre til periodens sekulariseringsprosesser og ikke bare til pressens utvikling. I en mediehistorisk fortolkningsramme kan imidlertid den gradvise økningen i «min mening»-kollokasjoner sees som et resultat av en ny kommunikasjonsteknologisk situasjon, der mennesket mot slutten av 1800-tallet fikk stadig flere plattformer og apparater for selv-eksteriorisering til disposisjon. I denne sammenhengen blir det dermed langt viktigere for oss å påpeke at den norske gullalderlitteraturen problematiserer hvordan fremveksten av et moderne mediasamfunn initierte former for subjektivitetsproduksjon som vi fortsatt arbeider med å håndtere i dag, enn å identifisere skillelinjen mellom romantikk og realisme.

Litteraturvitenskap i en digital tid

Digitaliseringen har endret våre lagrings- og kunnskapspraksiser. For en fremtidig litteraturvitenskap er det avgjørende å ta disse endringene på alvor, enten man omfavner digitale metoder eller ikke. I denne artikkelen har vi forsøkt å gå forsiktig frem for å utforske metodologiske tilnærminger tilpasset NBs grensesnitt. Motivert av mediehistorisk orienterte nærlesninger begynte vi med å spørre om NBs digitale arkiv kunne hjelpe oss med å utvide analysen av den moderne pressens påvirkning på litterære subjektivitetsuttrykk mot slutten av 1800-tallet. Gjennom prøving og feiling med noen av verktøyene som tilbys ved NB har vi løftet frem ny empiri som tyder på at den kommersielle pressens nye kommunikasjonsregime bidro til at et nytt selvforhold kommer til uttrykk. Slik peker utforskningen av det digitale arkivet i retning av at det sene 1800-tallets litteratur ikke bare representerer realismens «gjennombrudd» i litteraturen, men en begynnende forhandling av hva det vil si å leve i en moderne medietidsalder.

Hensikten med denne artikkelen har ikke vært å etterstrebe en uttømmende redegjørelse for digital litteraturvitenskap, men å gjøre noen innledende sonderinger for å utforske digitale lesestrategier i tilknytning til nb.no. Vår foreløpige konklusjon er, at det vil være galt ikke å oppdatere våre analysestrategier i tråd med at studieobjektene våre endrer seg. Dette betyr ikke at alle at hermeneutisk skolerte litteraturvitere må lære seg boolsk algebra eller programmeringsspråk, men at vår forskningspraksis kan berikes gjennom nye samarbeidsformer som DH kan gi oss.

Kollokasjonene som ligger til grunn for grafene ovenfor er radikalt annerledes studieobjekter enn det vi tradisjonelt har vært vant til i litteraturvitenskapen. Litteraturhistorisk sett mener vi at denne typen data kan bidra til både å styrke og utvide historiske undersøkelser, for eksempel slik vi har forsøkt å gjøre det i denne artikkelen. Litteraturteoretisk sett gjenstår det fortsatt å utforske hvordan DH påvirker litteraturvitenskapelige grunnspørsmål, og dette er ingen liten oppgave. Når boken og biblioteket så radikalt endrer materialitet, hva

gjør det med deres status som epistemologiske objekter? Digitaliseringen rokker ved grunnlagsspørsmål relatert til tekst, forfatter og kontekst og hvordan vi skal vurdere dem. For eksempel er det i dag kanskje vel så viktig å stille spørsmål ved hva en digital fil er og hvordan den digitale medieinfrastrukturen påvirker våre lesestrategier, som det er å undersøke det tradisjonelt definerende spørsmålet om hva en tekst er.

Litteraturvitenskapen fikk på godt og vondt en sentral rolle som teorileverandør til andre kulturvitenskaper opp gjennom forrige århundre. Med digitaliseringen er litteraturviteren nok en gang strategisk plassert til å kunne bidra med teoretiske og empiriske grunnpremisses til tilgrensende vitenskaper. En slik teoretisering ligger utenfor denne artikkelens rekkevidde, men det er verdt å merke seg at vi med Nasjonalbibliotekets satsning på digitalisering, som innebærer at Norge har et av verdens største og beste digitaliserte kulturhistoriske arkiver, burde ha gode forutsetninger for å sette premissene for fremtidens digitale litteraturvitenskap.

Litteratur

- Bastiansen, Henrik Grue. 2001. *Herom har jeg nærmere telegraferet til Kongen: telegraf og telefon i norsk politikk 1850-1940*. Oslo: Norsk telemuseum.
- og Hans Fredrik Dahl. 2008. *Norsk mediehistorie*. 2. utg. Oslo: Universitetsforlaget.
- Berry, David M. og Anders Fagerjord. 2017. *Digital humanities: knowledge and critique in a digital age*. Cambridge: Polity Press.
- Bjørnson, Bjørnstjerne. 1874. *Redaktøren: skuespil i fire handlinger*. København: Gyldendal.
- Bode, Katherine. 2012. *Reading by numbers: recalibrating the literary field*. London, New York: Anthem Press.
- . 2017. «The Equivalence of ‘Close’ And ‘Distant’ Reading; Or, toward a New Object for Data-Rich Literary History». *Modern Language Quarterly* 78 (1): 77–106. doi: [10.1215/00267929-3699787](https://doi.org/10.1215/00267929-3699787).
- Bok365. 01.02.2018 «Alt på plass på Bokhylla» Lest 26.06.2018. <https://bok365.no/artikkel/alt-pa-plass-pa-bokhylla/>
- Broadwell, Peter M. og Timothy R. Tangherlini. 2018: «Confusing the Modern Breakthrough: Naïve Bayes Classification of Authors and Works». *HumanIT* 14 (2): 19–42.
- Burdick, Anne, Johanna Drucker, Peter Lunenfeld, Todd Presner og Jeffrey Schnapp. 2012. *Digital humanities*. Cambridge, Mass: MIT Press.
- Cohen, Margaret. 1999. *The sentimental education of the novel*. Princeton, N.J.: Princeton University Press.
- Collett, Camilla. 1863. *I de lange Nætter*. Christiania: Cappelen.
- Dahl, Hans Fredrik, Kjell Lars Berge, Martin Eide, Nils E. Øy og Idar André Flo. 2010. *Norsk presses historie: 1-4 (1660-2010)*. Bind 1. Oslo: Universitetsforlaget.
- Dahl, Willy. 1981. *Norges litteratur 1 Tid og tekst 1814-1884*. Oslo: Aschehoug.
- Fitzpatrick, Kathleen. 2012. *The Humanities, Done Digitally*. Minneapolis, Minn.: University of Minnesota Press.
- Ganascia, Jean-Gabriel. 2015. «The Logic of the Big Data Turn in Digital Literary Studies». *Frontiers in Digital Humanities* 2 (7). doi: [10.3389/fdigh.2015.00007](https://doi.org/10.3389/fdigh.2015.00007).
- Golub, Koraljka, Marcelo Milrad, Marianne Ping Huang, Tolonen Mikko, Inés Matres og Andreas Bergsland. 2017. «Current Efforts, Perspectives and Challenges Related to Digital Humanities in Nordic countries» i International Symposium on Digital Humanities, Växjö, 7–8 November 2016: Book of Abstracts, pp.9–11. <https://open.lnu.se/index.php/isdh/article/view/422/370>
- Hamsun, Knut. 1893. *Redaktør Lynges Roman*. København: Philipsen.
- Hancher, Michael. 2016. «Re: Search and Close Reading». I *Debates in the digital humanities 2016*, redigert av Matthew K. Gold og Lauren F. Klein, 118–138. Minneapolis, Minn.: University of Minnesota Press.
- Hayles, N. Katherine. 2008. *Electronic literature: New Horizons For The Literary, Ward-Phillips lectures in English language and literature*. Notre Dame, Ind.: University of Notre Dame Press.

- Høyer, Svennik og Øyind Ihlen. 1995. «Forfattere i pressen. Noen data om forbindelesen mellom forfattere og pressen i det nittende århundret». *Norsk medietidsskrift* (1): 29–42.
- Jockers, Matthew L. 2013. *Macroanalysis: Digital Methods and Literary History, Topics in the digital humanities*. Urbana, Ill: University of Illinois Press.
- Kirschenbaum, Matthew. 2012. «What Is Digital Humanities and What's It Doing in English Departments?» I *Debates in the Digital Age*, redigert av Matthew Gold. Minneapolis, Minn.: University of Minnesota Press.
- Kittler, Friedrich A. 2009. *Mediefilosofi*. Oversatt av Knut Ove Eliassen. *Cappelens upopulære skrifter*. Oslo: Cappelen akademisk forlag.
- Lie, Jonas. 1884. *En malstrøm: fortælling*. København: Gyldendal.
- Malvik, Anders Skare. 2015. «The Advent of Noo-Politics in Ibsen's Problem Plays». *Ibsen Studies* 15 (1): 3–39. doi: [10.1080/15021866.2015.1087714](https://doi.org/10.1080/15021866.2015.1087714).
- Moretti, Franco. 2000a. «Conjectures on world literature.» *New Left Review* (1): 54–67.
- . 2000b. «The slaughterhouse of literature.» *Modern Language Quarterly* 61 (1): 207–228.
- NAOB. «kollokasjon». Lest 18.10.2018. <https://www.naob.no/ordbok/kollokasjon>
- . «min». Lest 18.10.2018. https://www.naob.no/ordbok/min_2
- Olsen, Jon Arild. 2014. «Når mer er bedre: Kvantitative litteraturstudier». *Norsk litteraturvitenskapelig tidsskrift* 17 (2): 139–145.
- Rinde, Harald. 2005. *Et telesystem tar form: 1855-1920*. Bind 1. Oslo: Gyldendal fakta.
- Schulze, H. 1866. *Skitser og Fortællinger*. Christiania: Dybwad.
- Seip, Jens Arup. 1971. *Ole Jacob Broch og hans samtid*. Oslo: Gyldendal.
- Søllinge, Jette D. og Niels Thomsen. 1989. *De danske aviser 1634-1989: Bind 2: 1848-1917*. Odense: Dagspressens Fond: I kommission hos Odense Universitetsforlag.
- Tangherlini, Timothy R. og Peter Leonard. 2013. «Trawling in the Sea of the Great Unread: Sub-corpus topic modeling and Humanities research». *Poetics* 41 (6): 725–749. doi: [10.1016/j.poetic.2013.08.002](https://doi.org/10.1016/j.poetic.2013.08.002).