

Grundtvigs Manuskripter Online

Maskinlæring og menneskeligt skøn



Af adjunkt, ph.d. Nanna Eva Nissen og lektor, lic.theol. Kim Arne Pedersen

N.F.S. Grundtvig (1783-1872) er ofte blevet kaldt den mest indflydelsesrige enkeltperson i Danmarkshistorien. Ved sin død efterlod han sig ikke blot et omfattende trykt forfatterskab, men også omkring 90.000 håndskrevne, upublicerede sider. Disse papirer opbevares i dag i N.F.S. Grundtvigs arkiv på Det Kongelige Bibliotek (KB).

Arkivet har længe været vanskeligt tilgængeligt – dels fordi opbygningen af det er uoverskuelig, dels fordi det kræver særlig træning at tyde Grundtvigs gotiske håndskrift. Derfor har forskningen kun i begrænset omfang udnyttet dette righoldige kildemateriale.

Det forsøger projektet *Grundtvigs Manuskripter Online* (GMO) at ændre, takket være bevillinger fra Augustinus Fonden og Aage og Johanne Louis-Hansens Fond i 2023. Over en fireårig periode bliver der opbygget en digital forskningsinfrastruktur i et samarbejde mellem KB og fakultetet, der skal gøre arkivet offentligt tilgængeligt i transskriberet og søgbar form.

GMO er planlagt og udføres af lektor Anders Holm fra fakultetet og specialkonsulent, cand.mag. Jakob K. Meile fra KB sammen med de to forfattere til

denne artikel. Projektet bygger på bibliotekets digitalisering af arkivet, og til at transskribere manuskripterne anvender vi Transkribus, en software til automatisk håndskriftgenkendelse.

Arkivets historie

Grundtvigs papirer var efter hans død i enken Asta Grundtvigs besiddelse. Ved hendes død i 1890 blev dokumenterne givet til Rigsarkivet og i 1941 overført til KB. Papirerne opbevares i dag under sikre forhold. I første omgang blev de bundtet i såkaldte fascikler og inden for de senere år ompakket i syrefri arkivæsker.

Inddelingen i fascikler er en nyttig struktur for det store, mangfoldige arkiv og er derfor fastholdt i ompakningen. I modsætning til de ordnede forhold, som arkivet i dag bevares under, er det interessant at tænke på, at dokumenterne fulgte Grundtvig gennem hele hans lange liv. Det vil sige lige fra han som 15-årig begyndte at gemme sit skriftlige arbejde, og da han siden flyttede det med sig mellem de ca. 25 forskellige adresser, han nåede at bo på. Man ved fra erindringer om hjemmet i Strandgade på Christianshavn i perioden 1828-40, at manuskripter lå spredt i et virvar sammen med bøger, blækhorn,

penne, tobakspiber og halvfulde te- og kaffekopper på et bord foran sofaen i arbejdsværelset.

Grundtvig giver selv i *Skolen for Livet* fra 1838 et levende indtryk af sin arbejdsplads – og af, hvordan ægtefællen Lise indimellem fik sat skik på den, når hun ryddede op og pudsede vinduer. På trods af sin umiddelbare modvilje og frygt for at miste orienteringen i sit velkendte kaos, måtte han indrømme nytten af rengøringen: “Alt hvad der før skulde sig i Skumringen under Støv og Papir-Dynger”, kom nu for dagens lys!

Arbejdsopgaver

Etableringen af forskningsinfrastrukturen hviler på to hovedopgaver, som varetages af ti studentermedhjælpere.

For det første skal alle affotograferede arkivdokumenter opmærkes, så hver linje og hvert ord placeres i korrekt læserækkefølge. Denne proces kaldes layouttilretning. Hvis et manuskript har en enkel struktur med overskrifter og afsnit, kan Transkribus-værktøjets automatiske linjeopmærkning ofte være tilstrækkelig. I vores tilfælde er manuskriptsidernes opbygning dog ofte så kompleks, at manuel justering af den automatiske opmærkning er nødvendig – især for at få Grundtvigs tilføjelser med, som tit er skrevet på kryds og tværs i margenerne. Dette arbejde kræver skarpe læseevner og dømmekraft i fastlæggelsen af den rette læserækkefølge.

For det andet skal der trænes en HTR-model (*handwritten text recognition*) i Transkribus, som kan maskintransskribere det opmærkede arkivmateriale. For at udvikle en model, der er specialiseret i at genkende Grundtvigs håndskrift, bruges et udvalg af manuelt transskriberede si-

der som træningsdata. I begyndelsen benyttede vi eksisterende transskriptioner, som blev matchet med de opmærkede billeder i Transkribus. De stammer fra Christian Thodbergs udgivelse af prædikenerne 1839-45, og fordi de følger et diplomatarisk princip og har alle rettelser med, er de velegnede til modeltræning i Transkribus.

Denne fremgangsmåde gav gode resultater. Med et træningssæt på ca. 1.100 sider eller ca. 290.000 ord havde vi tidligt i projektet en specialiseret model, der kan læse store dele af arkivet med en nøjagtighed på over 94 procent – og tekster fra 1840'erne helt op til 99,5 procent. Derudover fik studentermedhjælperne i denne indledende fase mulighed for at blive fortrolige med layouttilretningen og at læse sig ind på Grundtvigs håndskrift ved at støtte sig til de diplomatariske transskriptioner.

På nuværende tidspunkt handler det om at afprøve modellens formåen i andre genrer og årtier end de prædikener fra 1840'erne, som den hidtil har været trænet på. Her fungerer oplysninger om datering og indhold i den bindstærke registrant over papirerne fra Grundtvigsforskningens barndom som et vejledende styringsredskab, når vi prioriterer, hvilke dokumenter der skal transskriberes. Inden for disse prioriterede dokumenter har studentermedhjælperne nu mulighed for selv at vælge, hvad de ønsker at arbejde med. Deres opgave i denne fase består i at korrekturlæse HTR-modellens maskintransskription af dokumenterne og samtidig opmærke teksten med såkaldte *tags*.

Vi forventer at læse korrektur på ca. en tiendedel af arkivets sider for at kunne træne en model, der kan maskintrans-

skribere resten med en nøjagtighed på 96 procent eller derover.

Det daglige arbejde

Projektets studentermedhjælpere er ansat på KB og arbejder på Den Sorte Diamant, hvor de deler kontor med andre ansatte studenter tilknyttet bibliotekets specialsamlinger. Kontoret – kaldet ‘Transcriptoriet’ – ligger på fjerde sal med udsigt over Københavns havn og Knippelsbro. Udsigten giver mulighed for at løfte blikket og ‘kigge langt’ og fungerer dermed som et tiltrængt afbræk fra arbejdet med Grundtvigs håndskrift og Transkribus.

Det er nødvendigt at finde et godt *flow* i omgangen med manuskripterne. Ud over pauserne opnås variationen ved at skifte mellem layouttilretning og korrekturlæsning, som kræver forskellige former for opmærksomhed. Nogle dage fordyber man sig i sin egen skærm, andre dage opstår der livlige drøftelser på tværs af pc’erne, når man hjælpes ad med at knække vanskelige ord.

Mens dele af arbejdet er rutineprægede, kræver andre opgaver et skøn, som opøves over tid. Tydningen af Grundtvigs håndskrift er et godt eksempel: Målet er at læse teksten så bogstavret som muligt, men i praksis er det ofte nødvendigt at tyde ord ud fra, hvad der giver bedst mening i sammenhængen, fordi mange af Grundtvigs bogstaver ligner hinanden. Derfor er både erfaring med håndskriftlæsning og fortrolighed med Grundtvigs ordforråd vigtige forudsætninger for en kvalificeret tydning.

Med tiden udvikler man også en fornemmelse for, hvor meget energi man

skal lægge i at løse et vanskeligt layout eller afkode et uklart ord – og hvornår det er nødvendigt at lade det ligge og komme videre af hensyn til projektets fremdrift.

Linjen bagud

Den færdighed i læsning af Grundtvigs håndskrift og det kendskab til hans sprog, som GMO’s studentermedhjælpere oparbejder, erhvervede den første generation af Grundtvigforskere sig fra slutningen af 1930’erne og ind i de to første årtier efter 1945. Teologisk og humanistisk forskning handler også om at bringe tidligere tiders landvindinger ind i nutiden i opdateret og forbedret skikkelse, og GMO er ingen undtagelse. Digitaliseringen bygger videre på den ovennævnte *Registrant over Grundtvigs papirer*, som Grundtvigselskabet af 1947 sammen med Det Danske Sprog- og Litteraturselskab udarbejdede i 30 bind i årene 1954-64.

Registranten rummer beskrivelser af de enkelte fasciklers indhold og fungerer fortsat som et centralt redskab i arbejdet med GMO. Den er dog blevet til på baggrund af en langt mere ambitiøs plan om at gøre alle Grundtvigs tekster – trykte som utrykte – tilgængelige i en videnskabelig udgave. Planen blev fremsat af to af Grundtvigforskningens pionerer, litteraturhistorikerne Helge Toldberg og William Michelsen på opdrag af Hal Koch i 1946.

Betydning og potentiale

Når der blev iværksat en registrering frem for en udgivelse, skyldtes det, at Toldbergs og Michelsens udgave var beregnet til at omfatte 130 bind og derfor ikke havde gang på jorden. Nu er tiden en

anden: Den online-udgave af tekster publiceret i Grundtvigs levetid, som Center for Grundtvigforskning under AU gennemfører, vil sammen med GMO sikre, at næsten alt, hvad han har forfattet, bliver tilgængeligt.

Der tages derfor med GMO det sidste, store skridt til at kunne udforske Grundtvigs liv og værker. Det bliver muligt at søge på tværs af manuskripter og udgivne tekster, forskningen i Grundtvigs virke vil bevæge sig ad nye veje, og der vil blive kastet afgørende nyt lys over hans bidrag til Danmarkshistorien. GMO indebærer således hidtil usete muligheder for at undersøge Grundtvigs brug af kilder, da manuskripterne her er mere meddelssomme end de trykte tekster.

Allerede nu afslører forundersøgelser, at det hævdevundne billede af hans studier i oldtiden vil forandres, og at synet på hans forhold til det 18. århundredes videnskabelighed vil blive et andet. Videre vil GMO justere forskningen i Grundtvigs virke gennem det levende ord. Grundtvig talte frit, men tænkte altid sine præsentationer igennem i forudgående arbejdsmanuskripter. De er bevaret i arkivet, og GMO vil derfor ikke blot kunne opklare dunkle led i tidens ofte knappe referater af hans taler, men også belyse samtidige reaktioner på dem.

Dette er blot eksempler på GMO's videnskabelige afkast. Kun fremtiden kan vise, hvad der er i vente, men projektet vil uden tvivl ændre Grundtvigforskningen.

Fredag den 29. august 2025 kl. 14.00

Indvielse af Bodil Koch Auditoriet

Det Teologiske Fakultet markerer 150-året for kvinders adgang til universitetet med indvielsen af Bodil Koch Auditoriet

Sted: lok. 6B-0-22, Karen Blixens Plads 16, 2300 København S.

Arr.: Det Teologiske Fakultet