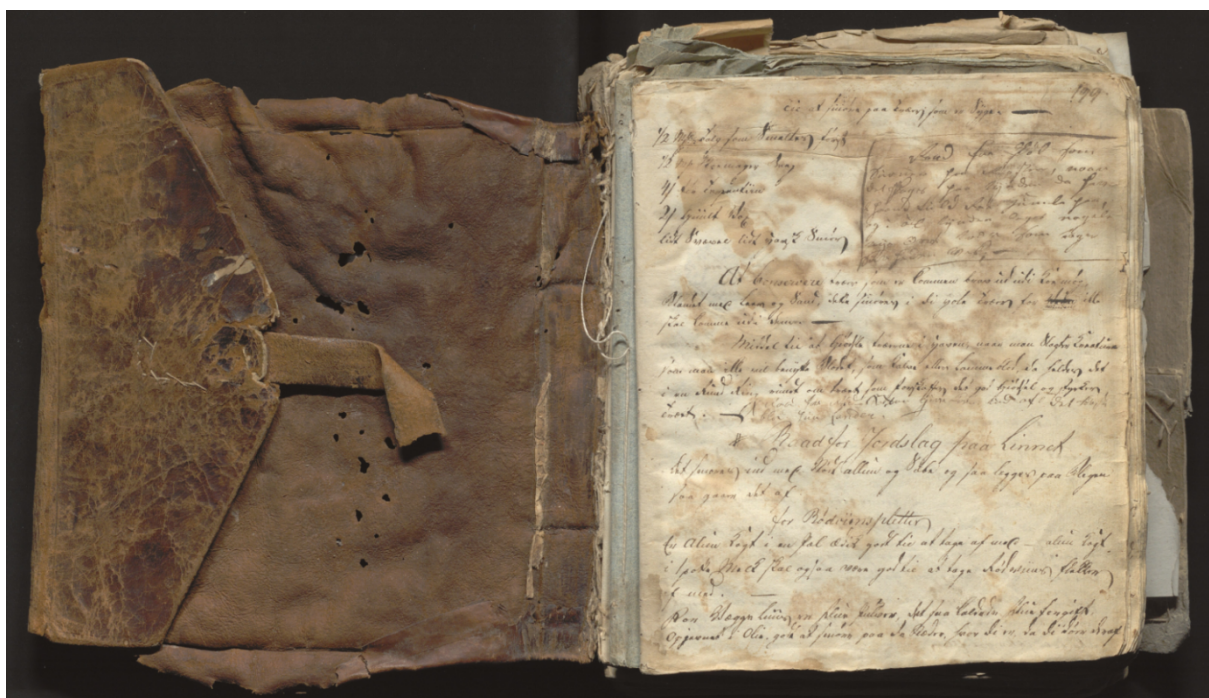


# Sluttrapport: Revitalisering av den kulinariske kulturarven: Tilgjengeliggjøring av norske kokebøker fra 1700- og 1800-tallet



Kildehenvisning: Utsnitt av Christine Storm Munchs kokebok, del 2 (Stiftelsen Lillehammer museum avd. Maihaugen)

## Innhold

Innledning.....	3
Mål og gjennomføring.....	3
Metode.....	5
Premiss for et godt resultat.....	5
Manuell transkribering og korrekturlesing.....	5
Prosess for automatisk tekstgjenkjenning med Transkribus.....	5
Erfaringer:.....	7
Konklusjon.....	8
Hvordan resultatene kan brukes videre.....	8

## Innledning

Innlandet fylkesarkiv har hatt et samarbeidsprosjekt sammen med Stiftelsen Lillehammer museum, avdeling Maihaugen, og det ideelle selskapet Tidvis. Prosjektets tittel er «Revitalisering av den kulinariske kulturarven: Tilgjengeliggjøring av norske kokebøker fra 1700- og 1800-tallet», muntlig beskrevet som kokebok-prosjektet.

Kokebok-prosjektet fikk støtte fra Arkivverket i 2021, på innsatsområde «Videreutvikling av privatarkivfeltet: Metodeutvikling, digital bevaring og tilgjengeliggjøring av arkiv». Prosjektet startet opp i løpet av 2022, men på grunn av mange ulike faktorer beskrevet i statusrapporten, ble prosjektet forsinket. Ny sluttdato ble satt til 01.12.2023.

## Mål og gjennomføring

Prosjektets overordnede mål er å etablere en god metodikk for å jobbe med kunstig intelligens i transkribering av eldre skrifter slik at de eldre kildene kan bli mer tilgjengelig for folk flest. Prosjektet har utgangspunkt i to andre prosjekter: *Food Lessons: Kulinarisk arv som ressurs i bygging av «Matnasjonen Norge 2030»* og det nordisk samarbeidet «*Mat – mer enn et måltid*». Kokebok-prosjektet føyer seg dermed inn i en større forskningssammenheng, ved at oppskriftssamlinger eller kokebøker blir digitaliserte og tilgjengeliggjort.

Opprinnelig var det 6 håndskrevne kokebøker som skulle inngå i prosjektet:

1. Christine Storm Munch kokebok, del 1 (Arkivverket/Statsarkivet på Hamar)
2. Christine Storm Munch kokebok, del 2 (Stiftelsen Lillehammer museum avd. Maihaugen)
3. Fruentimmer Kaage og Farvebog (Arkivverket/Statsarkivet i Stavanger)
4. Koke- og oppskriftsbok, gave fra L.D.Rogstad (Guneriusbiblioteket)
5. Walløes kokebok (i privat eie)
6. Loise A. Scholls kokebok (i privat eie)

De 4 sistnevnte kokebøkene var allerede avfotografert og i bruk av Tidvis. Derfor var det rom for å innlemme to andre kokebøker i prosjektet:

7. Ulrica Eleonora Fillion (1744/46-1806) (Stiftelsen Lillehammer museum avd. Maihaugen)
8. Kokebok for hvermand – af en Husmoder (i privat eie hos Julie Grindborg)

Dermed er 8 kokebøker avfotografert og tilgjengelige for forskning og formidling. Prosjektet kan imidlertid ikke kreve at andre tilgjengeliggjør kokebøkene i Digitalarkivet, og brukere må kontakte eierne for innsyn i de digitaliserte versjonene. Begge kokebøkene etter Christine Storm Munch og kokeboka etter Ulrica Eleonora Fillion er derimot publisert og tilgjengelige på Digitalarkivet.

Av disse 8 kokebøkene, er det de to kokebøkene til Christine Storm Munch som har dannet grunnlaget for transkribering med bruk av en Transkribus-AI innenfor dette prosjektets rammer. I forbindelse med bokprosjektet *Fru Christine Storm Munch og hennar kokebok* (Unn Bostad, Ivar Teigum, Arne Brimi og Torveig Dahl, Gudbrandsdalmusea 2016), har Unn Bostad fått transkribert store deler av den første delen av kokeboka til Christine Storm Munch (Arkivverket). Unn Bostad gav tillatelse til å bruke transkriberingen inn i prosjektet, i tillegg fikk prosjektet tillatelse fra Arkivverket/Statsarkivet på Hamar til å bruke protokollen.

I løpet av prosjektets tid har det vært mange utfordringer. Christine Storm Munchs kokebok del 2 måtte gjennomgåas nøye av konservatoren da det blant annet lå tørkede blomster i press mellom

## Sluttrapport

sidene, flere sider var festet med knappenåler, og noen sider var til og med limt med en klisterblanding. I tillegg var boka sydd sammen av mange forskjellige deler, og sydd slik at den var veldig skjør og flyttet lett på seg under arbeid. Dette medførte store tidkrevende problemer ved avfotografering i bokskanner. På denne måten brukte vi nesten 6 måneder på å få kokeboka klar til digitalisering og ferdig digitalisert. Kokeboka er registrert i Asta, publisert på Arkivportalen og ligger klar til bruk i Digitalarkivet.

Arkivportal/Christine Storm Munch, del 1: <https://www.arkivportalen.no/entity/bc8480aef5f-46f6-b9c0-8809ade51bcc?ins=AV>

Digitalarkivet/Christine Storm Munch, del 1: <https://www.digitalarkivet.no/source/70758>

Arkivportal/Christine Storm Munch, del 2: [https://www.arkivportalen.no/entity/no-OAM\\_arkiv\\_000000077425?ins=OAM](https://www.arkivportalen.no/entity/no-OAM_arkiv_000000077425?ins=OAM)

Digitalarkivet/Christine Storm Munch, del 2: <https://www.digitalarkivet.no/source/225260>

Arkivportal/Ulricka Eleonora Filion: [https://www.arkivportalen.no/entity/no-OAM\\_arkiv\\_000000077426?ins=OAM](https://www.arkivportalen.no/entity/no-OAM_arkiv_000000077426?ins=OAM)

Digitalarkiv/Ulricka Eleonora Filion: <https://www.digitalarkivet.no/source/225261>

Ifølge søknaden var den opprinnelige planen at Tidvis skulle holde kurs i Transkribus for Stiftelsen Lillehammer museum, avdeling Maihaugen og Innlandet fylkesarkiv. Nasjonalbiblioteket holdt et gratis innføringskurs i Transkribus, og en av Innlandet fylkesarkivs ansatte deltok på kurset. Tidvis holdt derfor en demonstrasjon for ansatte ved Stiftelsen Lillehammer museum og Innlandet fylkesarkiv /IKA Opplandene (KO). Demonstrasjonen ble gjennomført på Teams i begynnelsen av desember.

I august ble det bestemt at Stiftelsen Lillehammer museum, avdeling Maihaugen, skulle gi et stipend til en masterstudent tilknyttet Maihaugen, for at prosjektet kunne jobbe enda mer intensivt med Transkribus og på den måten få et bredere erfaringsgrunnlag. Oppgavene var: legge inn manus i Transkribus, layoutanalyse, teste forskjellige modeller for å innhente erfaringer, for å kunne utarbeide en god metode i arbeidet med Transkribus.

Innlandet fylkesarkiv har en gruppe med frivillige som har bidratt inn i prosjektet. Innlandet fylkesarkiv holdt kurs og opplæring i Transkribus for de frivillige, som opprettet egne brukere og har bidratt med følgende:

- Noen har transkribert sider manuelt
- Noen har gjort layoutanalyse, kjørt tekstgjenkjenning og gjort korrigeringer (korrekturlesning)
- Noen har gjort korrekturlesing på sider hvor layoutanalysen allerede har vært gjort
- En av våre frivillige skrevet historikk for kokebøkene til Christine Storm Munch.

I tillegg har Tidvis bidratt med profesjonelt transkriberte sier. Resten av kokeboka har vi brukt Transkribus til å transkribere. Dette resultatet ligger ved denne rapporten, og vil også bli publisert sammen med originalprotokollen i Digitalarkivet dersom det lar seg gjøre.

På vegne av prosjektet har Tidvis publisert en artikkel om Christine Storm Munch, samt 20 av hennes oppskrifter på lokalhistoriewiki:

[https://lokalhistoriewiki.no/wiki/Christine\\_Sophie\\_Storm\\_\(1746%E2%80%931825\)](https://lokalhistoriewiki.no/wiki/Christine_Sophie_Storm_(1746%E2%80%931825))

Prosjektet har hatt som overordnet målsetting om å etablere en god metodikk for å jobbe med kunstig intelligens i transkribering av eldre skrifter slik at de eldre kildene kan bli mer tilgjengelig for folk flest.

## Metode

### Premiss for et godt resultat

Transkribering forutsetter bruk av originalt kildemateriale, uansett om det er manuell transkribering eller ved hjelp av kunstig intelligens. Å bruke kunstig intelligens til transkribering innebærer at det originale kildematerialet er digitalisert med en tilstrekkelig høy oppløsning at Transkribus kan gjenkjenne tegnene eller bokstavene. Hvis oppløsningen på bildet er for lav, vil ikke Transkribus kunne identifisere bokstaver og tegn.

Valget mellom å bruke en offentlig modell, eller opprette en egen modell, bør avklares på et tidlig tidspunkt. I tillegg bør skriveregler og forkortelses avklares, da dette har stor betydning for å trene modellen.

### Manuell transkribering og korrekturlesing

I løpet av prosjektperioden ble 26 sider transkribert manuelt av to frivillige, 13 sider ble transkribert manuelt av ansatte på fylkesarkivet og 230 sider av ansatte i Tidvis. Innledningsvis ble det testet en rekke forskjellige offentlige modeller tilgjengelige i Transkribus, og den modellen som gav best treff på håndskriften i Christine Storm Munchs kokebok, var «19th Century Danish Gothic handwriting v.1.1». Denne modellen er brukt på ytterligere 70 sider som deretter er korrekturlest av frivillige og ansatte ved fylkesarkivet. I starten av prosjektet jobbet alle med layoutanalyse, før sidene ble kjørt gjennom modellen og til slutt korrekturlesing. På slutten av prosjektperioden ble resten av layout-analysen ferdiggjort av studenten, slik at det bare var korrekturlesing igjen for de frivillige.

### Prosess for automatisk tekstgjenkjenning med Transkribus

For å forberede de digitale filene for automatisk transkribering, må den igjennom en forprosess, hvor det første steget var å få lastet opp filene i Transkribus. Neste steg var å gjøre layoutanalyse på alle sidene. Layoutanalyse innebærer å tegne opp alle felt med tekst og tekstlinjer, for å fortelle Transkribus hvor den skal finne teksten den skal analysere.

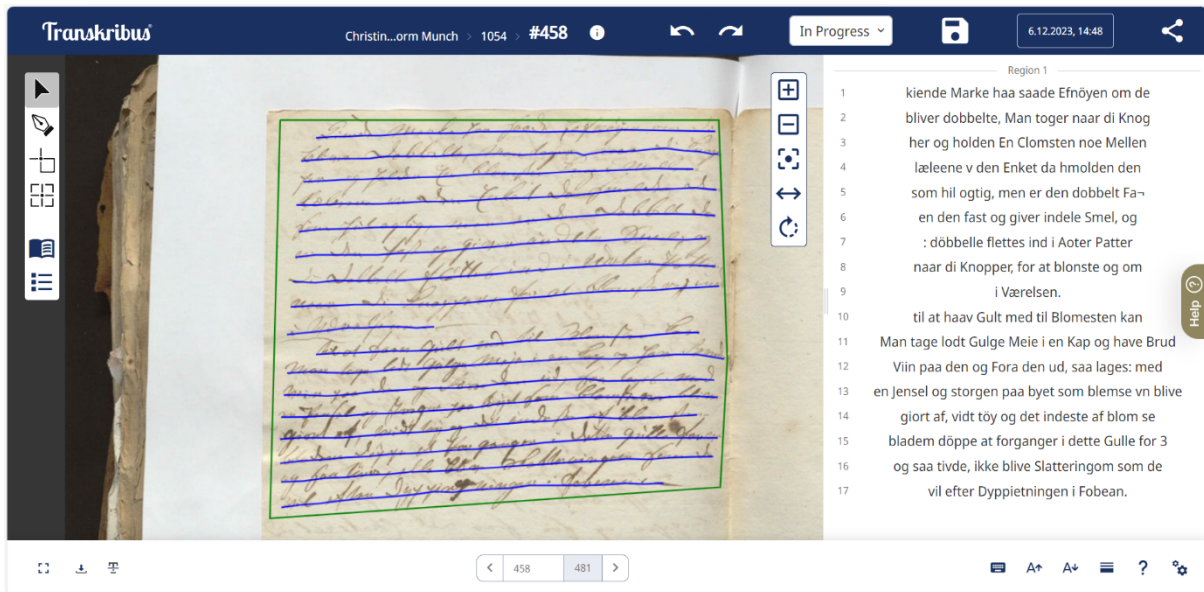
Det finnes en innebygget versjon av layoutanalyse i Transkribus. Den fungerer godt, men det må gjøres en del korrigeringer i etterkant. Den får ikke alltid med seg for eksempel slutten på linjer, hopper over noe tekst, eller deler opp teksten på det som for oss er ulogiske steder.

Neste steg var å kjøre tekstgjenkjenning i Transkribus. I programvaren ligger det en del håndskriftsmodeller inne. De modellene som er offentlige i Transkribus er laget av institusjoner som har jobbet på forskjellige prosjekter, ofte med fokus på en spesiell tidsperiode eller type håndskrift. Den modellen som viste seg å fungere best på Christine Storm Munchs kokebok er «19th Century Danish Gothic handwriting v.1.1». Link: <https://readcoop.eu/model/danish-gothic-handwriting/>

Modellen har en feilprosent på 6,7 %, som innebærer at den tar feil på 6-7 bokstaver per 100 ord på tilsvarende type materiale som den er trent på. Det er vanskelig å bedømme hvor godt den treffer akkurat på Christine Storm Munchs kokebok, da lesbarhet ikke bare går på prosentvis feil, men også hvordan vi subjektivt oppfatter en tekst. Eksempelet under viser at man kan forstå hva teksten

## Sluttrapport

handler om, men om man for eksempel skal søke etter ord i den digitale teksten, så kan det være for mye lesefeil fra Transkribus til at man får treff.

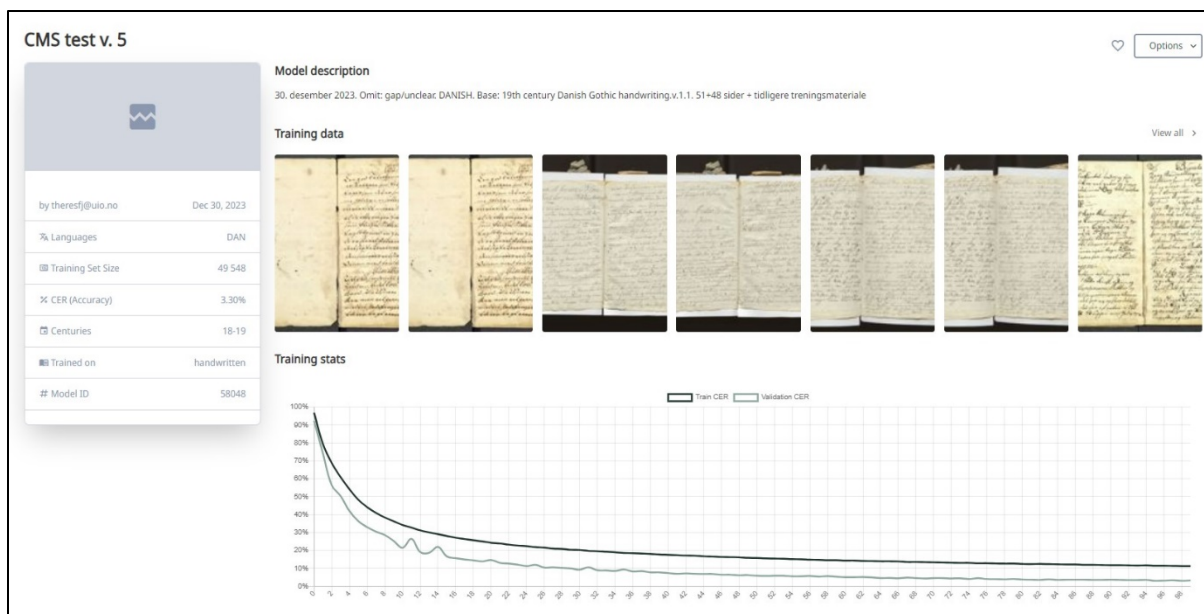


Figur 1: Venstre side viser utsnitt av en side av kokeboka, området vi har definert som tekst (grønn rute) og linjene (blå) som Transkribus har gjenkjent. Høyre side viser Transkribus sin transkripsjon

Håndskriften som brukes i Christine Storm Munchs kokebok er fra perioden 1750-1850, og per nå ikke finnes en god nok modell som passer på en så variert håndskrift. Med «god nok» menes en automatisk tekstgjenkjenning som ikke krever noe etterarbeid, men som kan publiseres som det er. Alle modeller som ble testet ut krever en del etterarbeid i form av korrigering av teksten. Det er noe tidkrevende, men likevel går det raskere enn å transkribere manuelt. Den valgte modellen gir en god lesestøtte til store deler av teksten, og klarer tolke en del håndskrift korrekt nok til at det viser innholdet i kokeboka.

Underveis høsten 2023, ble det utarbeidet fasitsider, såkalte Ground Truth- sider, som ble brukt til å teste ut å lage en egen modell. Modellen per 31. desember, ved prosjektets slutt, inneholder 48 filer fra Christine Storm Munchs kokebok del 1, og 51 filer fra Christine Storm Munchs kokebok del 2, pluss noen duplikater fra et par tidligere tester. Den er bygget på modellen «19th Century Danish Gothic handwriting v.1.1», som vil si at ny modell bygger videre på statistiske gjetninger av hvilke ord eller bokstaver den leser, basert på grunnmodellen. Feilraten er på 3,30 % på de testene som blir gjort av Transkribus under modelltreningen. Testene gjøres på samme type materiale som er brukt til treningen. Transkribus anbefaler at man trener minst 20 sider på den håndskriften man skal transkribere. Christine Storm Munchs kokebok inneholder mange forskjellige håndskrifter, og dermed bør man generere fasitsider, så godt det er mulig, fra alle typer håndskrift som er i boka for at den skal dekke godt nok. Likevel lover det godt at de testene som er utført denne høsten viser en lav feilrate.

Modellen ble deretter testet på noen sider i boka, og en subjektiv vurdering av teksten som ble generert er at den er lesbar, men foreløpig ikke vesentlig bedre enn «19th Century Danish Gothic handwriting v.1.1».



Figur 2: Visning av modellen i Transkribus, med antall ord den er trent på og eksempler fra materialet. Grafen viser at etter hvert som Transkribus har jobbet med treningsmaterialet, så har treffprosenten på bokstaver blitt bedre og bedre.

Det som ble transkribert fra bunnen av ble det lagt veldig mye tid i for å få det så korrekt som mulig, slik at det kunne benyttes som Ground Truth senere. Det viste seg at for de som ikke er profesjonelle vil det på de fleste oppslag være ett eller annet tegn som er utydelig, kanskje på grunn av stavefeil. I nyeste oppgradering av Transkribus kan man likevel holde unna de ordene som er usikre, og likevel lage Ground Truth av resten av siden. Det vil kunne øke tilfanget av Ground Truth-sider betraktelig.

I kokeboka finnes det mange forkortelser som gjelder mål og vekt, og en av utfordringene er hvordan Transkribus skal transkribere tegn som ikke finnes i vårt alfabet, men som skal læres av modellen (Ground Truth). Ved å være konsekvent i bruken av det i Ground Truth, så vil modellen kunne kjenne igjen og lære nye tegn. Eksempel ble tegnet for pund (libra): ꝛ testet ut i egen modell, og den ble gjenkjent.

## Erfaringer:

Ettersom det er en langsom prosess å transkribere for hånd, var det et ønske om å teste ny teknologi i arbeidet med å transkribere håndskrevne manus for å undersøke om bruk av kunstig intelligens kunne gjøre dette arbeidet mindre tidkrevende, for på denne måten gjøre håndskrevne kilder mer tilgjengelig for folk flest.

Å arbeide med kunstig intelligens utfordrer tankegangen og etablerte rutiner for transkribering. Tidligere var det teksten og bokstavene som var i sentrum, og man begynte gjerne et transkriberingsarbeid fra første side og jobbet seg gjennom protokollen. Med kunstig intelligens er det selve verktøyet og mulighetene verktøyet kan tilrettelegge for som blir det viktigste. Ettersom det er skriftgjenkjenningen som er viktig, blir et utvalg sider gjennom hele protokollen viktig for å kunne trene modellen.

Selve forarbeidet, med avfotografering og layout-analyse, er en tidkrevende prosess. Imidlertid er det nettopp forarbeidet som er det aller viktigste å få på plass for å lykkes i prosjekter som dette. Videre er det valg av modeller og skriveregler som har innvirkning på transkripsjonens kvalitet.

## Sluttrapport

En god layout-analyse forutsetter at en er tett på originalmaterialet. Dette for å kunne forstå oppbyggingen av teksten, ordvalg og tegnsetting. Å finne setningene på de mest rotete sidene i kokeboka, må man ha transkribert ordene først, før man vet i hvilken rekkefølge de/linjene skal stå. I disse tilfellene er faktisk den manuelle transkriberingen viktigst.

De frivillige som deltok i prosjektet, syntes det var spennende å jobbe med transkribering på en ny måte. Alle savnet en god nybegynnerveiledning, for det viste seg vanskelig å lære seg Transkribus uten et godt opplæringsverktøy. Innlandet fylkesarkiv holdt derfor samlinger for de frivillige, slik at de ble trygge i arbeidet. Det var særlig layout-analysen som viste seg vanskelig å forstå/utføre i Transkribus. Prosjektets student fikk dermed ansvaret for å gjøre layout-analysen, mens de frivillige fikk ansvaret for å korrekturlese transkripsjonen og foreta korrigeringer. Dette var en veldig effektiv arbeidsmåte.

## Konklusjon

I dag finnes det ingen håndskriftsmodell som er god nok til at man kan bruke den uten korrigeringer. Foreløpig er det for mange feil til å kunne bruke transkripsjonene generert i Transkribus som noe annet enn en enkel lesestøtte til de digitale filene. I tillegg tar det lang tid å gjøre transkripsjonene fra Transkribus lesbare. Det viser at det er et fortsatt behov for å utvikle modeller som kan brukes på norsk materiale fra denne perioden som Christine Storm Munchs kokebok er skrevet i. Prosjektet har lagt et grunnlag for videre arbeid med håndskriftsmodeller. Generelt sett blir modellene bedre, dess mer materiale den mates med.

For de som er profesjonelle eller godt vant til å transkribere for hånd, går det raskere eller like raskt manuelt som å gå via Transkribus. For de som er bare middels gode i transkribering, er det lettere å korrekturlese enn å transkribere alt manuelt. Disse vil kunne spare litt tid på å kjøre tekstgjenkjenning først. Vår erfaring er at når layout-analyse først var gjort, gikk transkriberingen enda raskere.

### Hvordan resultatene kan brukes videre

De sidene som til nå er fasitsider, vil bli liggende i Transkribus og kan brukes videre inn i andre prosjekter eller samarbeid. Aarhus City Archives, som står bak språkmodellen «19th century Danish Gothic handwriting v.1.1», er interessert i å snakke om et samarbeid. De skal på sikt oppdatere modellen sin, og er interessert i å se om transkripsjonene som er generert gjennom kokeboksprosjektet, kan inkluderes i modelltreningen.

Det er opprettet kontakt med både Nasjonalbiblioteket og prosjektet SAMLA – digitalisering av norske tradisjonsarkiv (<https://samla.w.uib.no/>), der vi har snakket om å lage en felles håndskriftsmodell for denne tidsperioden. Det er per nå ingen konkrete planer om oppstart.

I løpet av våren 2024 vil det i masterprosjektet som er tilknyttet kokebokprosjektet, jobbes videre med korrigering av teksten og utvikling av modell/metode. Målet er at det skal være en god nok modell som kan brukes på tilsvarende materiale/periode for andre prosjekter.