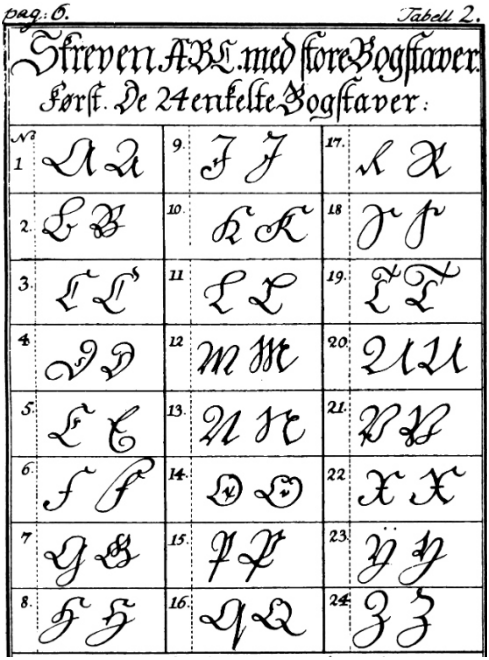
Den glemte skrift

Om gotisk håndskrift, historiske kilder og AI-baseret tekstgenkendelse

Af Anne Ingeborg Sørensen

»Krøllede bogstaver« er det populære udtryk for en skrift, der fra reformationstiden og indtil slutningen af 1800-tallet var den dominerende i Danmark. Den trykte version hedder fraktur, mens håndskriften kaldes *gotisk håndskrift* eller *nygotisk kursiv*. Skriften opstod i Nordfrankrig og Flandern i 1100-tallet. At den blev kaldt gotisk, var oprindeligt en nedsættende betegnelse, der stammer fra de italienske renæssancehumanister. De havde med forlæg i de klassiske romerske store bogstaver (versaler/majuskler) udviklet den enkle, runde og flydende trykskrift antikva, og den afledte håndskrift blev kaldt latinsk eller humanistisk skrift.


pag. 6. Tabell 2.

Skreven ABC. med store Bogstaver.

Først. De 24 enkelte Bogstaver.

Figur 1. Fra *Skreven Dansk ABC*, Kbh. 1761. Her gengivet efter Georg Simon: *Gotisk skrift*. Kilde: slaegtsbibliotek.dk.

De kantede bogstaver, der blev anvendt længere nordpå, fik navn efter folkeslaget goterne, der hærgede i Romerriget flere århundreder før og blev opfattet som barbariske og primitive.

Begge håndskrifter er udviklet fra det latinske alfabet, så skellet mellem dem er på sin vis kunstigt. Især i Tyskland og Skandinavien tog man efterhånden betegnelsen *gotisk skrift* til sig, og den blev opfattet som mere ægte, nordisk/germansk og folkelig end den »fremmede« latinske skrift antikva. I England og i de romanske lande brugte man i langt højere grad antikva på tryk og latinsk håndskrift.

I løbet af 1800-tallet kom der i Danmark kritik af den gotiske kursivskrift, i første omgang blandt intellektuelle. Det spredte sig til skoleverdenen, hvor børnene efterhånden skulle lære skrive- og læsefærdighed i både latinske og gotiske bogstaver for at kunne læse bøger og lære sprog fra andre lande end

Tyskland. Sverige brugte fx i langt højere grad antikva og latinske bogstaver i tryk og i håndskrift på grund af tættere forbindelser til Frankrig. Det var desuden i både tryk- og håndskrift almindeligt, at fremmedord samt person- og stednavne stod med latinske bogstaver.

At børnene skulle lære »dobbelt skrift« var genstand for pædagogiske skoledebatter fra midten af 1800-tallet. N. F. S. Grundtvig kaldte i *Berlingske Tidende* (der var trykt i fraktur på det tidspunkt) forslagene om at afskaffe gotisk skrift for en »Grusomhed mod gamle, troe Tjenere« (17. december 1846). Hans begrundelse var, at det dels ville gøre flere århundreders litteratur utilgængelig for mange, dels at det ville være svært for den almene befolkning at opnå færdighederne. Vi er her i en tid, der nok havde en befolkning med en vis generel læsefærdighed, men først efter indførelsen af skoleloven i 1814 blev skrivefærdighed alment udbredt. Og især på landet var der fortsat mange, der kun kunne læse og skrive til husbehov.

Grundtvigs ønske blev dog ikke opfyldt. Ved et cirkulære i 1875 blev det bestemt, at man i skriveundervisningen »for Fremtiden indskrænker sig til det ene Alphabet, hvilket da maa være det latinske«. Baggrunden var pædagogisk, men det har sikkert også spillet ind, at gotisk skrift blev associeret med tysk sprog og kultur, og at der efter nederlaget i 1864 og afståelsen af Slesvig-Holsten var kommet en anti-tysk stemning i Danmark. Det tog dog en generations tid, før den gotiske håndskrift helt forsvandt – og i Nordslesvig, der indtil 1920 var tysk, var gotisk skrift fortsat den almindelige. I 1943 konstaterede historikeren Erik Kroman i bogen *Skriftens Historie i Danmark*, at »det er nu kun faa af den ældre Generation, der kan læse, endsige skrive den gamle danske Skrift« (s. 64).

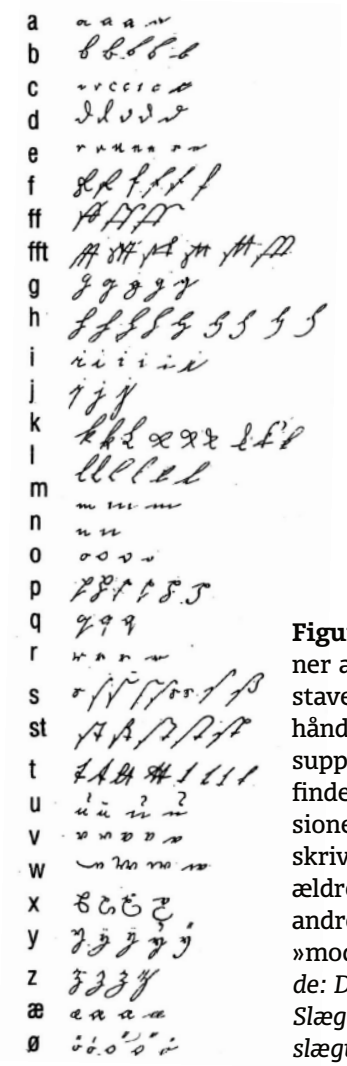
At læse og tyde gotisk håndskrift kræver i dag øvelse og erfaring. Skriften er på mange områder markant anderledes end den latinske, og for mange er den i dag stort set ulæselig. Slægtsforskere og historikere er traditionelt de grupper, der har opøvet kompetencer i at læse den, fordi de har brug for at kunne tyde håndskrevne kilder fra før 1900. Men der er ikke længere obligatorisk undervisning i håndskrift på historieuddannelserne, og mange har derudover også problemer med at afkode den latinske håndskrift, der er udbredt i 1900-tallets kilder og materiale. Håndskrevne kilder er derfor efterhånden blevet uforståelige for mange og bliver derfor ikke anvendt.

Kendetegn og fælder. Når *e* ligner *n*, og *h* ligner *f*

Skriveredskaber var afgørende for både skriftens udbredelse og dens afskaffelse. Man brugte oprindeligt gåse- eller svanefjer, hvor spidsen var skåret til på tværs. Det gav mulighed for at lave brede streger, når pennen blev ført nedad, og tynde streger, når den blev ført til siden. Derfor er de gotiske bogstaver kendetegnet ved kantethed, spidse vinkler og mange krøller og sløjfer. For at undgå for mange blækklatter var det også bedst at undgå for mange stop i skriveforløbet, så nogle bogstaver fik en udformning, der støttede flowet. Det gør sig blandt andet gældende ved lille *s* og lille *h*, der for en nutidig betragtning begge ligner en slags *f*-er. Mange bogstaver kan i det hele taget let forveksles (se figur 2), så det kan være svært at afgøre, om der står *c* eller *i*, *e* eller *n*, *r* eller *v* eller *f*, *h* eller *s*. Det kan være næsten umuligt at se forskel på *m*, *n*, *e* og *u*. Skriften er desuden kendetegnet ved mange ligaturer (sammenskrivninger), fx af *s* og *t*, dobbelt-*s* og dobbelt-*f*.

Skriften blev efterhånden mere flydende og sammenhængende og mindre stejl. Det hang sammen med flere ting. Dels fik håndskriftens kommunikative funktion forrang over dens tidligere mere æstetiske, kunstfærdige og til tider ophøjede/religiøse værdi. Papir blev billigere, og administration og embedsværk fyldte mere og mere i de fremvoksende nationalstater. Det var derfor en fordel, at det gik hurtigere. Håndstillingen ændrede sig, så håndfladen hvilede på papiret, pennen lå mere skråt i stedet for at pege opad, og man kunne derfor bedre skrive uden at løfte pennen, der desuden var blevet mere spids og satte tyndere og mere ensartede streger.

I løbet af 1800-tallet vandt fabriksfremstillede stålpenne udbredelse. De var tyndere og mere stive end fjerpennene, og de krævede ikke det samme tryk på papiret. Det gav udfordringer i forhold til den gotiske håndskrift, der havde mange knæk og vinkler. Det skabte faldgruber i form af risikoen for blækklatter, mens stålpennen var optimal til den latinske skrift, der var baseret på runde, jævne og lette strøg. Overgangen fra gotisk til latinsk håndskrift var derfor også teknologisk dre-



Figur 2. Versioner af små bogstaver i gotisk håndskrift. I supplikkerne findes alle versioner, da nogle skrivere brugte ældre bogstaver, andre var mere »moderne«. *Kilde: Danske Slægtsforskere, slægt.dk.*

vet og ikke kun begrundet i pædagogiske/praktiske hensyn og ideologi.

Ny teknologi: fra papir over mikrofilm til digitalisering

Ny teknologi er også årsagen til, at kendskabet til de mange kilder, der er skrevet i gotisk håndskrift, i de seneste år har vundet større udbredelse, fordi det med computerhjælp er blevet muligt at digitalisere og tyde skriften. For nogle årtier siden begyndte man at fotografere udvalgte historiske kilder (især arkivalier og aviser) og gøre dem tilgængelige på mikrofilm. Det betød en hurtigere arbejdsgang end at skulle bestille arkivkasser til brug på læsesale, men det var en omstændelig proces at sidde foran mikrofilmapparaturet med halvdårligt baggrundslys og spole filmene igennem for at gennemgå fx avisårgange for relevante artikler om et bestemt emne. Ofte valgte forskere på grund af umuligheden ved en komplet søgning i store komplekse samlinger at lave systematiske stikprøver ved at læse kilder fra bestemte år, måneder eller dage.

Med digitalisering af materialer er OCR (Optical Character Recognition/optisk tekstgenkendelse) blevet et værktøj. Nu kan trykte og scannede sider transformeres til søgbar tekst, og det giver hurtigere overblik over større tekstmængder, fordi man med »distant reading« kan bruge metoder, der anvender computere til analyser (»fjernlæsning«) af store tekstmængder. Og målrettede ordsøgninger kan på samme vis hurtigt finde relevante emner i et stort tekstkorpus. Det er den teknik, der ligger bag mange digitale kildeudgivelser på nettet, fx på Rigsarkivet, Det Kongelige Bibliotek og i litterære digitale værksamlinger.

Kilder i håndskrift har dog ikke været tilgængelige på samme måde, fordi der er meget store udsving i skrivemåder og de enkelte skriveres kendetegn. De seneste år har kunstig intelligens været teknologien bag sprogmodeller, der kan trænes til at afkode og transskribere håndskrift ved at genkende og kategorisere bogstaver og samtidig søge at sætte ord ind i en for den pågældende tekst plausibel kontekst. Man styrer selv, hvilket materiale man vil træne sin specifikke model på, så der er altså tale om en tilrettet og specifik sprogmodel til forskel fra de mere generelle AI-sprogmodeller som ChatGPT og andre LLM'er (Large Language Models). Det er dog ikke perfekte transskriberinger, man ender med. Som ved andre AI-modeller forekommer det, at modellen »hallucinerer« og gætter på sandsynlige ord, hvis der er udfordringer med at afkode bogstaverne. Det ender derfor nogle gange i rent volapyk, som man løbende må arbejde på at begrænse ved at gentræne og optimere modellerne.

Samlet set giver de nye sprogmodeller dog mulighed for at bruge de ældre kilder meget mere end før – og for at bruge dem på nye måder. Man kan søge efter navne, emner, steder etc., og man kan også bruge et tekstkorpus til undersøgelser, hvor digitale metoder fx kan analysere, hvilke emner der er fremtrædende i bestemte perioder, hvilket ordvalg der bruges, hyppige navneord, udviklinger i sprogbrugen over tid, samt hvilke geografiske lokaliteter, befolkningsgrupper og fænomener der er mest fremtrædende – eller mest fraværende.

***Voices of the People*: 1700-tallets supplikker og embedsværk vakt til live**

Et af de nye AI-værktøjer til tekstgenkendelse er det europæiske Transkribus, der efterhånden har fået stor udbredelse i arkiv- og forskningsverdenen til transskribering af ældre hånd- og trykskrifter. Historikere ved Aarhus Universitet gør i projektet *Voices of the People* enevældens supplikprotokoller fra 1699 til 1799 tilgængelige og digitalt søgbare ved brug af Transkribus. Supplikker er bønskrifter, klager eller ansøgninger, som befolkningen i det dansk-norske monarki havde ret til at sende til den enevældige konge. »Kongen« udvikler sig i løbet af denne periode til i mange tilfælde at være centraladministrationen og embedsværket, der voksede og fordrede sagsbehandling, ensrettethed, bureaukratisering og arkivering. Kongen selv fik forelagt særlige sager til afgørelse, fx om benådning for henrettelse, mens kancelliernes embedsmænd sagsbehandlede og afgjorde de mere gængse ansøgninger.

Supplikprotokollerne indeholder ikke de originale supplikker, som i de fleste tilfælde er sendt tilbage til ansøgerne sammen med svaret eller er blevet kasseret, men afskrifterne gengiver ofte de centrale dele og argumenter fra den originale supplik. Supplikprotokollerne er således både en kilde til befolkningen og dens livsverden, til dens interaktion med det administrative og politiske system samt til embedsværkets og enevældens normer og ageren. Alle befolkningsgrupper skrev til kongen: mænd og kvinder, fattige og rige, udlændinge og folk fra hele det danske rige, som på det tidspunkt indbefattede Norge, Slesvig-Holsten, Island, Færøerne, Grønland samt kolonier og handelsstationer i Vestindien, Afrika og Indien. Det er især bemærkelsesværdigt, at fattige, almindelige bønder og håndværkere samt kvinder figurerer som indsendere; det er befolkningsgrupper, som der ellers er meget få kilder fra i denne periode.

Befolkningen sendte ansøgninger om personlige forhold som tilladelse til ægteskab mellem beslægtede, for at få såkaldt »kongelig konfirmation« på testamenter og uskiftet bo og for at få en stadfæstet afgørelse i konflikter om fx penge, ejendele eller rettigheder. De skrev for at få udsat, fremskyndet eller sløjfet en verserende retssag og for at få fri proces eller strafnedsættelse. Mange ansøgninger handlede også om at opnå erhvervs- og forsørgelsesmæssige privilegier, fx tilladelse til at drive handel eller få bevilget en pension. Enevældens embedsmænd i hele riget indsendte desuden supplikker med konkrete forslag til samfundsforbedringer, fx i forbindelse med fattighjælp, vejanlæg, epidemier og brandsikkerhed.

Det har hidtil krævet særlige kompetencer og masser af tid at få overblik over supplikprotokollernes indhold.

Originalerne befinder sig på Rigsarkivet, da det er ENB-materiale (af »enestående national betydning«), og man kan ikke få lov at se originalerne. Rigsarkivet har scannet protokollerne, men det er et korpus på ca. 225.000 tætskrevne sider, og der er kun sporadisk mulighed for gennem den originale og rudimentære indholdsfortegnelse med ansøgernes navne at slå op i protokollerne.

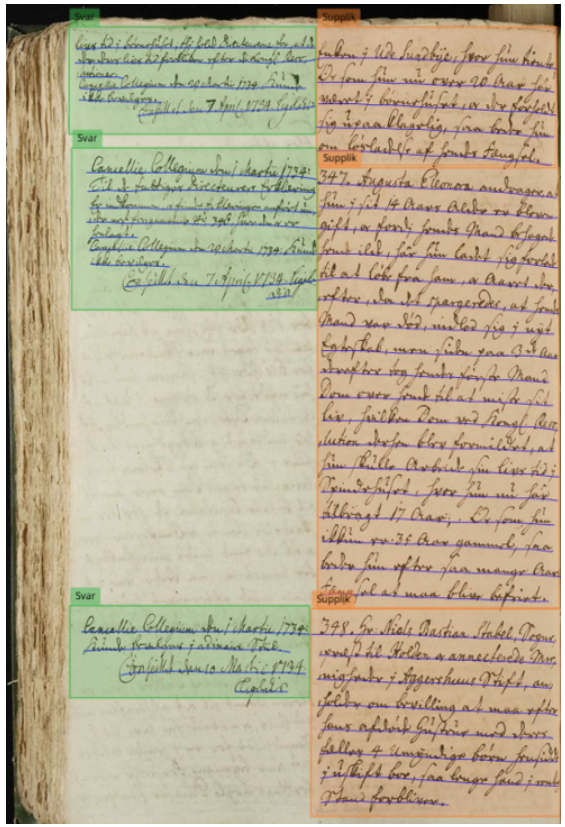
Ved hjælp af Transkribus og digitaliseringen er der nu mulighed for at søge systematisk på personer (fx Tordenskjold og Struensee), emner (fx drab, fødsler i dølgsmål, luksusvarer, håndværk), steder (fx provinsbyer, Grønland og Dansk Vestindien), befolkningsgrupper (fx enker, jøder, slaver, syge) og embedsværkets ageren (fx epidemier, skat, import).

Et søgeresultat er dog ikke nødvendigvis perfekt. Vi har trænet en model, som i lighed med andre Transkribus-modeller til trykte og håndskrevne kilder på forskellige sprog og fra forskellige perioder er offentligt tilgængelig og kan bruges af andre. Den er løbende blevet trænet på det materiale, vi har bearbejdet og læst korrektur på. Det vil sige, at den kender mange »hænder« (skrivere) og mange ord. Vi er efterhånden nået ned på, at modellen kun »læser forkert« i ca. 3 % af materialet, hvilket er rigtig godt. Men ethvert søgeresultat skal tages med forbehold. Der kan dukke en ny skriver op, der har en anden måde at skrive bestemte bogstaver på, og det forvirrer modellen. Nogle skrivere er også nogle rodehoveder, der ser stort på spalter og konventioner eller skriver meget gnidret og småt.

Sproget i supplikkerne

Som sproglig kilde giver supplikkerne et indblik i 1700-tallets hverdag, samfund og magtforhold. Da de som nævnt kun findes i referater, udfærdiget af embedsmændene, har vi ikke mulighed for at se, fx hvordan en fæstebonde formulerede sig, når han klagede over en ridefoged. Ofte fik man også en skriver – en såkaldt koncipist – til mod betaling at udforme skrivelsen, der så fremstod mere formfuldendt og korrekt. Især i begyndelsen af 1700-tallet nævnes det af og til, at ansøgerne har underskrevet med deres forbogstaver, hvilket betyder, at de næppe kunne læse og skrive. Man kan dog ane »almindelige menneskers« sprog i mange supplikker, når de skal redegøre for et sagsforløb eller en konflikt. Til gengæld er supplikkerne en fremragende kilde til embedsværkets og enevældens sprogbrug. Det er en stil, der holder af lange sætninger, semikoloner, få punktummer, et virvar af indskudte sætninger og mange fremmedord. Det er ikke uden grund, at det kaldes kancellisprog – supplikkerne blev sagsbehandlet i Danske Kancelli, der var det øverste centrale forvaltnings- og regeringsorgan. Det er dog tydeligt, at der er forskel på skrivernes evner og/eller forkærlighed for denne stil, for der findes også eksempler på mere nøgterne og kortfattede supplikker og svar.

Sprogbrugen er ofte fremmed for en nulevende dansker. Mange ord er mere eller mindre uddøde, som *lejermål* (sex



Figur 3. Foto af supplikprotokol 1734-1, s. 184 med transskribering til højre.

I udsnittet af protokolsiden fra 1734 ses det, hvordan vores layoutanalyse af den håndskrevne tekst udføres i Transkribus. Siden køres gennem en field-model, der analyserer strukturer og kategoriserer dem som enten supplik (lysebrun baggrund) eller svar (grøn baggrund). Derefter køres en layout-model, der registrerer linjer og deres begyndelse og slutning (blå linjer). Dette arbejde kaldes segmentering og er i vidt omfang automatiseret, men bliver kontrolleret af projektets ansatte (historikere og studentermedhjælpere).

Efter segmenteringen køres tekstmodellen på materialet, og den bliver tilføjet de enkelte opslag i den digitale protokol, så brugeren kan se både originalen og tekstmodellens version. Den transskriberede tekst giver et indblik i et samfund med streng moral og hårde domme. Den ansøgende kvinde fik et klart afslag på sin ansøgning om at komme ud af spindehuset (en straffe- og arbejdsanstalt i København) efter en livstidsdom for bigami. Spargeredes *betyder, at det blev fortalt/det rygtedes.*

uden for ægteskabet), *hjemgæld* (selvforskyldt tab eller skade), *rodemester* (skatteopkræver), en *hører* (underlærer) og *gangklæder* (hverdagstøj). Andre ord har ændret betydning, som fx *billig* (berettiget, rimelig) og *uopdragen* (ikke voksen, uforsørget). I arbejdet med at forstå sprogbrugen er *Ordbog over det danske Sprog* (1700–1950) et uvurderligt redskab, fordi oplysninger om både forskellige stavemåder og betydninger er medtaget i dette værk.

Nogle gange er ordene mere arkaiske eller specielle og må slås op i *Holbergordbogen*, *Renæssanceordbogen* (1500–1700), *Kalkars Ordbog* (1300–1700) eller *Gammeldansk Ordbog* (1100–1515). Ofte må man også på jagt i fremmedsprogene. Den danske enevældes administration var på især det juridiske område præget af andre europæiske landes konventioner og principper, så mange ord og begreber skrives på fransk, latin eller tysk. Nogle eksempler er *insultere* (fornærme), *acquiescere* (indvillige i, acceptere), *emanere* (udstede) og *beneficium paupertatis* (fri proces grundet fattigdom).

Før ordbøger og officielle retstavningsregler var stavning af ord desuden meget skiftende. Det var afhængigt af den enkelte skrivers baggrund og kunnen, og mange ord blev gengivet ud fra, hvordan de blev udtalt, fx mange person- og stednavne. Konsistens var heller ikke altid noget, man lagde vægt på, så i den samme tekst kan et navn sagtens være stavet på flere forskellige måder. Endelig var sproget dengang som nu under konstant udvikling, så nogle skrev *ck* istedet for *kk* i ord som *ikke* og *c* i stedet for *g* i *og*, mens andre hurtigere hoppede på de mere danske og moderne udgaver. Det kan give udfordringer for sprogmodellen. Vores model vil gerne please og komme med et bud på et ord, den ikke rigtig kan finde ud af, fordi den ikke genkender bestemte skriftkendetegn eller stavningen, og så griber den ned i skuffen med meget brugte ord som *Ægteskab*, *Enke* og *Stervboe* (dødsbo), som er hyppigt forekommende, fordi testamenter og tilladelser til uskiftet bo skulle omkring enevældens administration. Modellen har også lidt svært ved årstal, fordi den oprindelig er »fodret« med mange data fra 1758-protokollen og derfor vender tilbage til det årstal, når der opstår udfordringer med aflæsningen.

Hvad med fremtiden?

AI og tekstmodellerne har gjort mange støvede arkivalier tilgængelige for både forskere og offentligheden. De er efterhånden også så gode, at de er et brugbart værktøj, især til at følge historiske emner over en længere tidsrække gennem målrettede ordsøgninger. Søgningerne er dog aldrig perfekte, og der kan være ting, der ikke fanges i modellerne. Som eksemplerne i illustrationerne viser, er transskriberingen heller ikke nødvendigvis letforståelig for en nutidig læser, fordi datidens sprogbrug ikke altid er umiddelbart genkendelig. Nogle forskere leger med tanken om at udvikle AI-værktøjer, der kan automatisere en versionering af teksten til fx moderne dansk. Der er også forudsigelser om, at de mere traditionelle AI-modeller/

LLM'er (ChatGPT, Gemini, Claude etc.) inden for en overskuelig tid kan gøre det samme som et specialtrænet program som Transkribus, men vi er ikke nået dertil endnu. Problemet er blandt andet, at de ikke er trænet på ens specifikke materiale, så de bliver næppe gode til 1700-tallets danske administrationssprog inden for den nærmeste fremtid. Desuden sikrer programmer som Transkribus også en viden om, hvad modellen er trænet på. Man ejer fortsat sine egne data og bestemmer selv, om og i hvilket omfang det må bruges af andre. Det er ikke i samme grad tilfældet med LLM'er.

Der er ingen tvivl om, at modellerne tilgængeliggør sproglige kilder, der for historikere og offentligheden kan belyse ting, der hidtil har været skjult eller glemt. Det er dog også vigtigt at være opmærksom på, at de større projekter, der p.t. er i gang, i høj grad er finansieret af eksterne bevillinger og fonde, og at det dermed er kurateret, hvilke samlinger og kildekorpusser der når offentligheden. Det forhindrer dog ikke, at alle kan udforske en platform som Transkribus (se link nedenfor) ved at bruge en gratiskonto til fx at uploade tipoldemors breve eller sider fra den gamle familiebibel i fraktur og finde ud af, hvad der egentlig står.

Anne Ingeborg Sørensen,
historiker, ph.d., daglig projektleder af
Voices of the People,
Afdeling for Historie og Klassiske Studier,
Aarhus Universitet

Links

Voices of the People, *Aarhus Universitet: https://cas.au.dk/voices. Projektleder: Professor Nina Koefoed. Projektet bliver realiseret med en bevilling fra Carlsbergfondet.*

»Tekst-model, 1700-tallets administrative gotiske håndskrift, september 2025« *(projektets tekstmodel): https://app.transkribus.org/models/text/397597*

Transkribus: *https://www.transkribus.org/da. Man kan oprette en gratis konto til upload af et begrænset antal filer.*

Hvis man vil lære at læse gotisk håndskrift, har Rigsarkivet et fint onlinekursus: https://haandskrift.slaegt.dk/