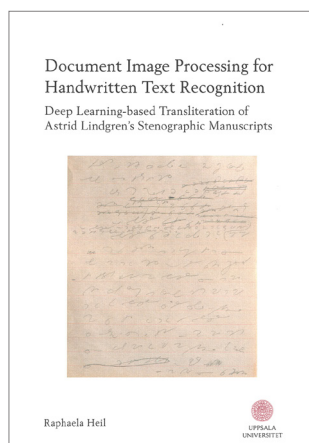


Review/Recension

RAPHAELA HEIL  
**DOCUMENT IMAGE  
PROCESSING FOR  
HANDWRITTEN TEXT  
RECOGNITION**  
**Deep Learning-based  
Transliteration of Astrid  
Lindgren's Stenographic  
Manuscripts**



Uppsala universitet, 2023. Skrifter utgivna av Svenska barnboksinstitutet nr 166 (87 + 75 s.)

Raphaela Heils avhandling *Document Image Processing for Handwritten Text Recognition. Deep Learning-based Transliteration of Astrid Lindgren's Stenographic Manuscripts* (2023) närmar sig Astrid Lindgrens författarskap från ett perspektiv ur vilket det kanske aldrig tidigare har betraktats. Avhandlingen är skriven vid Institutionen för informationsteknologi vid Uppsala universitet, men arbetet har också till stora delar varit knutet till projektet *Astrid Lindgren-koden*. Detta projekt har dragit samman litteraturvetenskaplig forskning, datavetare och professionella stenografer som tillsammans tolkat Lindgrens stenogram med hjälp av digitala metoder. Det är just denna tvärvetenskapliga och digitala miljö som Heils datavetenskapliga avhandling vuxit fram inom.

Utöver det uppenbara litteraturvetenskapliga och historiska värdet av att tillgängliggöra en central källa och förstå ett viktigt författarskap, finns det också mer datavetenskapliga anledningar att ta sig an just denna uppgift. Det senaste decenniet har det skett en häpnadsväckande utveckling av metoder för datorbaserad tolkning av handskrivna text, *Handwritten Text Recognition* (HTR). Från att tidigare endast ha varit ett område för informationsteknologisk grundforskning har denna teknik på senare år blivit en del både av humanistisk forskning och kulturarvsinstitutioners insatser för digitalt bevarande och tillgängliggörande av historiskt material.

©2025 Jacob Orrje. This is an open access article distributed under the terms of the Creative Commons CC BY 4.0 License (<https://creativecommons.org/licenses/by/4.0/>), permitting all use, distribution, and reproduction in any medium, provided the original work is properly cited. Any included images may be published under different terms. Please see image captions for copyright details. Citation: *Barnboken - tidskrift för barnlitteraturforskning/Barnboken: Journal of Children's Literature Research*, Vol. 48, 2025 <http://dx.doi.org/10.14811/clr.v48.985>

Teknikens framväxt är tätt knuten till 2000-talets revolution inom maskininlärning och kanske mer specifikt djupinlärning (på engelska *Deep Learning*) – det vill säga algoritmer som använder sig av så kallade neurala nätverk med flera lager för att lära datorer att generalisera mönster i stora datamängder. Sådana algoritmer, som mer förutsättningslöst etablerar samband genom att träna på så kallade dataset (i detta fall exempelvis matchade par av å ena sidan bilder av ord och å andra sidan transkriberad text av samma ord) har på senare år i många fall ersatt tidigare regelbaserade system (det vill säga program där människor på förhand via programkod i detalj har beskrivit exempelvis hur en viss bokstav ser ut). Inom textigenkänning har dessa datadrivna metoder möjliggjort en exceptionell breddning av mängden tillämpningar. Där det för drygt ett decennium sedan enbart fanns programvaror för breda kommersiellt gångbara uppgifter (exempelvis transkribering av modern tryckt text), är det idag möjligt att hitta system för att, mer eller mindre träffsäkert, transkribera bland annat frakturtryck, historisk handskrift och icke-västerländska skriftsystem.

Man kan säga att Heils avhandling befinner sig i fronten av arbetet med att förbättra datorers förmåga att tolka ovanligare historisk skrift. Lindgrens stenogram är i detta sammanhang av flera skäl en särskilt svår uppgift och avhandlingen syftar huvudsakligen till att utveckla lösningar för denna utmaning. För det första innehåller manuskripten en mängd redigeringar i form av överstrykningar, ändringar och inskjutningar. Detta problem, att tolka förändrade manuskript där text delvis är dold, har inte utforskats i någon större utsträckning av tidigare forskning inom HTR. För det andra utgör själva det stenografiska skriftsystemet – i Lindgrens fall Melins system, det mest kända svenska systemet för stenografi – en utmaning för HTR-tekniken. Innan projektet *Astrid Lindgren-koden* fanns det en utbredd uppfattning att Lindgrens manuskript var nästan omöjliga att avkoda. Svårigheten för otränade läsare att tillgodogöra sig texten kommer från det faktum att stenografin är baserad på en större mängd tecken än vanlig skrivstil samt att skriften är mer kompakt där mindre skillnader i pennföringen kan vara betydelsebärande. Dessa svårigheter skapar även tekniska utmaningar och ett andra spår i avhandlingen, utöver att hantera redigeringar i form av överstrykningar, är att utveckla lösningar för att lära datorer att läsa Melins system. Som Heil påpekar är hennes uppgift mer komplicerad än att enbart utveckla metoder för att transkribera manuskripten. Hennes mål är nämligen en digital metod för att translitterera

materialet, det vill säga en lösning för att tolka de stenografiska tecknen och översätta dem till vanlig text.

Nu är det här en teknisk avhandling och det ingår mer eller mindre i genren att texten bitvis är skriven på ett sätt som kanske inte tilltalar gemene man eller den mer humanistiskt skolade forskaren. Detaljerade matematiska formler, redogörelser för olika lager av neurala nätverk och avancerade tekniska redogörelser talar snarare inåt mot det datavetenskapliga fältet än till en bredare publik eller tvärvetenskaplig läsekrets. Även om det begränsar möjligheten att tillgodogöra sig forskningen bland forskare och kulturarvsinstitutioner som arbetar med dessa frågor är denna tekniska stil knappast något som går att invända mot – denna typ av inomvetenskapliga tilltal är ett resultat av de krav som alla akademiska ämnen ställer på sina respektive doktorander. Samtidigt är texten ändå ofta skriven på ett relativt pedagogiskt sätt, som gör det möjligt att följa författaren allteftersom hon närmar sig och sedermera löser problem. Bakgrunden (kapitel 2) innehåller föredömligt tydliga begreppsdefinitioner inte bara av tekniska termer utan av själva studieobjektet. Exempelvis: hur kan vi kategorisera olika delar av ett ord på en sida, vad är stenografi, vad skiljer en translitteration från en transkription, vad är en diplomatisk transkription och hur är Melins stenografiska system uppbyggt? På samma sätt diskuteras centrala tekniska begrepp inom bildanalys (så som digitala bilder, binarisering och morfologiska operationer) samt inom maskininlärning och djupinlärning på sätt som även ger något mindre datavetenskapligt skolade läsare en möjlighet att hänga med.

De följande två kapitlen presenterar de vetenskapliga papers som avhandlingen bygger på. Kapitel 3 fokuserar på de olika överstrykningar som kan dölja ord i manuskript – exempelvis enkla och dubbla linjer, enkla och dubbla diagonala linjer, vågiga linjer eller kludd där linjer går fram och tillbaka horisontellt över ett ord flera gånger. För att med hjälp av maskininlärningsmetoder kunna lära datorer att identifiera och ta bort dessa typer av överstrykningar skapar Heil ett antal dataset med parvisa matchande ordbilder – det vill säga par av identiska ordbilder men där den ena bilden innehåller en överstrykning och den andra saknar sådan. En fråga som diskuteras i kapitlet är hur ett sådant dataset kan byggas upp. Å ena sidan kan man manuellt skanna sidor utan överstrykningar, för att sedan stryka över ord på sidorna och skanna dem igen. På så sätt får du digitala bilder av verkliga överstrykningar, vilket kan möjliggöra träning på data som mer direkt liknar framtida tillämpningar. Samtidigt räcker

det inte för djupinlärningsalgoritmerna att du har skannat samma sida, utan bildparen måste passas mot varandra så att pixlarna av orden matchar mot varandra (annars riskerar datorn att lära sig fel saker om skillnaderna mellan bilderna när den övar på dem). Att i efterhand matcha bilderna så att själva orden ligger identiskt på pixelnivå kräver mycket manuellt arbete och gör det alltså svårt att sammanställa de stora datamängderna som krävs. Ett alternativt tillvägagångssätt som avhandlingen utforskar är att skapa syntetiska överstrykningar – som alltså läggs till av datorn efter att ordet är skannat. För att kunna göra sådana datorgenererade annoteringar studerar Heil därför hur materialet ser ut i detalj: hur långa är överstrykningarna i manuskripten, hur breda är de, var i höjddled på orden förekommer de, vilken textur har de och vilken färgton? Utifrån denna analys bygger hon en teknisk lösning som skapar syntetiska överstrykningar som liknar verkliga så mycket som möjligt. Denna automatiska lösning är såklart mer arbetseffektiv, men Heil påpekar samtidigt hur den riskerar att göra så att datorn tränar på data som inte är lika verklighetsnära. När problemet och möjliga lösningar väl är definierade får vi följa med författaren när hon låter olika djupinlärningsalgoritmer träna på data skapad på dessa olika sätt, för att slutligen utvärdera resultatet.

På samma sätt diskuterar kapitel 4 hur HTR-teknik kan användas för att tolka stenografi. Här presenteras LION-datasetet som innehåller delar av Lindgrens manuskript skrivna i Melins system – huvudsakligen "Bröderna Lejonhjärta" men också utdrag ur "Emil i Lönneberga" och delar av exempelvis biografiska och självbiografiska texter. Kapitlets data bygger på den crowdsourcing som genomförts inom ramen för projektet *Astrid Lindgren-koden*, där kunniga stenografer har translittererat delar av Lindgrens manuskript (sammanlagt 198 sidor och 2900 linjer). Kapitlet visar att denna stenografiska text är mer svårarbetad för dagens HTR-teknologi än dataset med vanlig handstil av motsvarande storlek, men att sådan teknik ändå kan tolka handskrivna stenografi till viss del. Vi ser också att träffsäkerheten kan förbättras genom en kombination av tekniska lösningar (förträning) och att koda träningsdata på ett sätt som tar hänsyn till de speciella stenografiska tecken som finns i Melins system (så som förkortningar samt speciella tecken för inledningar och avslutningar av ord). På ett liknande sätt som i det föregående kapitlet visar Heil här igen prov på en god förmåga att kombinera tekniskt kunnande med detaljkunskap om det skriftsystem som

ska bearbetas – här alltså stenografisystemets specifika karaktär. Kanske ser vi här en effekt av att Heil verkat inom ett projekt som inte bara möjliggjort tvärvetenskaplig samverkan mellan forskare med litteraturvetenskaplig och datavetenskaplig kompetens utan som också dragit till sig personer utan koppling till universitetsvärlden med djup praktisk sakkunskap. Det är intressant att se att denna typ av kunskapsmässig korsbefruktning, som blir allt vanligare inom så kallad digital humaniora, också ser ut att gynna datavetenskaplig forskning.

Heil nämner i avhandlingen att det varit svårt att få tillräckligt många volontärer genom crowdsourcing för att skapa ett större annoterat dataset, huvudsakligen på grund av hon genomfört sitt avhandlingsprojekt under Covid-pandemin. Detta knyter an till det kanske enda större frågetecken som dök upp under min läsning av avhandlingen. Jämfört med många andra dataset för HTR-tolkning som byggts upp de senaste åren, exempelvis de med historisk handskrift (där det dataset som ligger till grund för Riksarkivets snarlikt döpta modell "The Swedish Lion" i ett svenskt sammanhang är det största på omkring 16 miljoner ord), är storleken på den data som avhandlingen analyserar relativt liten. Eftersom kvaliteten på utfallet av just den typen av djupinlärningsmetoder som diskuteras i avhandlingens fjärde kapitel är starkt beroende av större mängder data, går det alltså att anta att tolkningen av stenografin hade kunnat förbättras påtagligt bara genom att utöka mängden av data som algoritmerna fått träna på. Här hade det varit intressant med en mer ingående analys och uppskattning av hur träffsäkerheten hos de automatiska translitterationerna berodde på den underliggande tekniska arkitekturen och hur mycket som helt enkelt snarare skulle kunna avhjälpas med mer data. Hur stort problem hade exempelvis överstrykningar varit vid en större mängd träningsdata? Hade det varit nödvändigt att anpassa kodningen av data till Melins system om vi hade haft träningsdata på säg 1000 sidor i stället?

Samtidigt är det å andra sidan kanske just detta fokus, på att göra det mesta utifrån ett mindre dataset genom att verkligen förstå detaljerna i de stenogram som ingår i det, som i mina ögon är avhandlingsarbetets stora styrka. Denna tvärvetenskapliga ansats som kombinerar datavetenskaplig expertis med ett genuint intresse för materialet, är en styrka och skapar en intressantare inblick i materialet än vi hade fått exempelvis av utvärderingar av träffsäkerheten i modeller tränade på olika stora mängder data. Denna tvärvetenskapliga ansats gör även att läsare utan expertis inom det direkta

datavetenskapliga fält inom vilket avhandlingen är skriven kan läsa den med behållning: både för att få en inblick i hur transformerande teknologier så som maskin- och djupinlärning fungerar i praktiken, men också för att få inspiration för hur de kan användas i exempelvis ett historiskt eller litteraturvetenskapligt sammanhang.

*Jacob Orrje  
Docent i idéhistoria  
Uppsala universitet*