

ERFAHRUNGS- UND ERGEBNISBERICHT ZUM TRANSKRIBUS-PROJEKT:

*»ps: ich bitt noch mahl umb ver gebung meines
confusen und üblen schreibens wegen«*

Frühneuzeitliche Briefe als Herausforderung automatisierter Handschriftenerkennung

Martin Prell

Jena, Mai 2018

Mail: [martin.prell\[at\]uni-jena.de](mailto:martin.prell[at]uni-jena.de)

Twitter: [@prellmartin](https://twitter.com/prellmartin)

Inhalt

1. Projektgegenstand und -ziele.....	1
2. Quellenspezifische Herausforderungen	1
3. Projektumsetzung	3
3.1. Konzeption und Leitung.....	3
3.2. Trainingsdatenerstellung (Ground-Truth)	3
3.3. Projektspezifische und -übergreifende Aspekte, Fragen und Entscheidungen	4
3.3.1. Automatische vs. händische Layoutanalyse	4
3.3.2. Buchstabensuspension am Zeilenende	4
3.3.3. Zeilenwechsel vor Blattende	6
3.3.4. Unterscheidung kurrenter und lateinischer Buchstaben	6
3.3.5. Überschriebene Wörter und Buchstaben	7
3.3.6. Haupthand und andere Schreibhände	8
3.3.7. Einbezug von Wörterbüchern	8
3.3.8. Offline-Arbeiten.....	9
3.3.9. Abstand der Baseline zum Zeichen.....	9
3.3.10. Automatisierte Zuordnung von Text- und Bildzeilen.....	10
4. Projektergebnisse und deren Weiterverwendung.....	10
5. Subjektive Einschätzungen und Ausblicke	13
6. Übersicht Projektdaten	13

1. Projektgegenstand und -ziele

Am Historischen Institut / Lehrstuhl für Geschlechtergeschichte der Universität Jena wurde im Oktober 2017 ein Projekt begonnen, das die Absicht verfolgte, eine Technologie zu erproben, um zwei Forschungsschwerpunkte des Lehrstuhls in unmittelbarer Anwendung aufeinander weiter zusammenzuführen: die **Digital-History-Forschung** und die Untersuchung von **(weiblicher) Schriftpraxis frühneuzeitlicher Selbstzeugnisse** (<http://www.histinst.uni-jena.de/Bereiche/Geschlechtergeschichte/Projekte/Automatische+Handschriftenerkennung.html>). Bei der eingesetzten Technologie handelt es sich um „**Transkribus**“ (<https://transkribus.eu>), eine Software zur automatisierten Erkennung und Transkription chirographischer Texte, die im Rahmen des e-Infrastruktur-Projektes READ (Recognition and Enrichment of Archival Documents: <https://read.transkribus.eu/>) entwickelt wird. Bereits im September 2016 erfolgte in Jena ein Workshop zur Anwendung von Transkribus mit dem Titel „Automatisierte Text- und Strukturerkennung als Grundlagentechnologie für die Digital Humanities“ (siehe dazu den Workshopbericht in H-Soz-Kult vom 17.11.2016, www.hsozkult.de/conferencereport/id/tagungsberichte-6821).

Mit dieser Software steht – neben der Geschichtswissenschaft – allen historisch arbeitenden Textwissenschaften bereits seit geraumer Zeit ein computerbasiertes Werkzeug zu Verfügung, dessen Anwendung auf (nicht nur aber insbesondere) handschriftliche Quellen neue Dimensionen historischen Forschens verspricht. Die Software entfaltet ihr Potential insbesondere dort, wo Handschriften in großem Umfang, sauberem Schriftbild, gleichmäßiger Ausführung und guter Qualität von Schriftträger und digitalem Faksimile vorhanden sind. Dazu gehören insbesondere serielle Quellen aber beispielsweise auch Korrespondenzkonvolute eines/einer Schreibers/in. Zu letzteren sind auch die **Briefe der pietistischen Gräfin Erdmuthe Benigna von Reuß-Ebersdorf (1670-1732)** aus dem frühen 18. Jahrhundert zu zählen, die 2017 in einer digitalen Edition herausgegeben wurden (URL: <http://erdmuthe.thulb.uni-jena.de/erdmuthe-digitale-edition/start.html>).

Während der zweijährigen Projektlaufzeit der Edition konnte nur ein Teil der Gesamtkorrespondenz der Gräfin auf hohem historisch-kritischem Niveau ediert werden (66 Briefe, ca. 250 Seiten). Der überwiegende Teil der bis dato bekannten Korrespondenz dieser Frau und Vorfahrin zahlreicher europäischer Königshäuser musste zunächst unbearbeitet bleiben. Es handelt sich dabei um über 100 weitere Briefe mit ca. 500 Seiten. Damit waren zwei wichtige **Kriterien** erfüllt, die einen zeit- und kosteneffizienten Einsatz der Transkribussoftware versprochen: Auf der einen Seite existierten bereits Transkriptionen in elektronischer Form und ausreichendem Umfang (ca. 50-100 Seiten werden empfohlen), die sich als Ground-Truth-Daten zum Trainieren eines speziell für die Briefe zu erstellenden Recognition-Modells eigneten. Auf der anderen Seite war eine ‚kritische Masse‘ an noch untranskribierten Briefen vorhanden, die den mit dem Einsatz der Software einhergehenden Zeit-, Arbeits- und Kostenaufwand gegenüber einer konventionellen ‚händischen‘ Transkription rechtfertigte.

Das Projekt war auf zunächst 2 ½ Monate angelegt. Es hatte dafür personelle Mittel im Umfang von zwei wissenschaftlichen Hilfskraft-Stellen (à 40 h/Monat), die ausschließlich in diesem Projekt ange stellt waren, sowie einer wissenschaftlichen Mitarbeiter-Stelle, die das Projekt als Ergänzung zu einem anderen Digital-Humanities-Projekt des Lehrstuhls („Editionsportal Thüringen“: <http://www.histinst.uni-jena.de/Bereiche/Geschlechtergeschichte/Projekte/Editionsportal+Th%C3%BCrtingen.html>) ausführte, zur Verfügung.

2. Quellenspezifische Herausforderungen

Ein zentrales Anliegen des Projektes bestand im gezielten Austesten der derzeitigen **Möglichkeiten und Grenzen der Transkribus-Software**, die in anderen Anwendungsfällen bereits hervorragende Er-

gebnisse erzielt hat. Für gotische Buchhandschriften ließ sich mit Material des Editionsprojektes Parzival eine Zeichenfehlerquote unter 3% erreichen (13. und 14. Jh.). Die Hände der Sekretäre von Jeremy Bentham sind mit 5% Character Error Rat (CER) erkennbar (18. Jh.). Protokollbände in regelmäßiger Kurrent der Schweiz lassen sich aufbauend auf einem generellen Modell ohne Trainingsdaten mit unter 10% CER identifizieren.

Die Briefe der Gräfin stellen insofern eine **besondere Herausforderung** für die computerbasierte Erkennung dar, als sie mit ihrem selbst für frühneuzeitliche Verhältnisse geringen Normierungsgrad, ihrer im Vergleich mit ausgebildeten Schreibern unkonventionellen Graphetik und sehr unregelmäßigen Orthographie hilfreiche Kriterien für eine korrekte automatisierte Erkennung scheinbar gerade nicht bedienen. Damit repräsentieren die Briefe aber pars pro toto einen wichtigen Umstand der Realität historischer Quellenbestände, der eine Grundherausforderung von HTR-Software schlechthin darstellt, nämlich die letztlich unbegrenzte Heterogenität der Erscheinung handgeschriebener Dokumente. Handschriften zeichnen sich – im Gegensatz zu Druckschriften – gerade durch ihr **hohes Maß an freier Gestaltungsmöglichkeit** aus, wobei die chirographisch stark regulierten Buchschriften des Mittelalters hier einmal ausgenommen und der Blick vielmehr auf die Geschäftsschrift gerichtet sei. Diese Heterogenität betrifft letztlich alle Aspekte einer handgeschriebenen Quelle: die Ausführung der kurrenten Buchstaben und deren kursive Verbindung zu Wörtern, die räumliche Nutzung und Platzierung des Schriftbereiches, nachträgliche Änderungen und Hinzufügungen, der zugrundeliegende Schreibstoff und die verwendeten Schreibgeräte etc. All diese Aspekte zusammengenommen machen jedes handgeschriebene Dokument und dessen Set an Gestaltungsmerkmalen einzigartig, was die auf möglichst wiederkehrende Strukturen und Abtrennbarkeit von Buchstaben angewiesenen OCR-Verfahren schnell überfordert.

Die andere Herausforderung ist sozio-historischer Art und betrifft – vor allem in Bezug auf weibliche Schreibpraxis – auch Syntax und Rechtschreibung des Geschriebenen. Eine höhere Geistesbildung und die damit einhergehende Unterweisung im gelehrten Schreiben war für Frauen im frühen 18. Jahrhundert noch eine absolute Ausnahme. Selbst innerhalb adliger Kreise und deren ausgebildeten Erziehungssystem war dies in der Regel ein Privileg männlicher Zeitgenossen, die damit auf eine zwangsläufig bevorstehende oder potentiell mögliche Übernahme der Regierungstätigkeit vorbereitet wurden. Lesen und Schreiben waren hier nicht allein intellektuelle Kulturtechniken, sondern schlichtweg zwingend erforderliche Notwendigkeiten zur Aufrechterhaltung und Fortsetzung der Dynastie. Das bedeutet aber nicht, dass weibliche Adlige zwangsläufig Analphabetinnen waren. Ihnen kamen sogar meist ebenfalls dynastische Aufgaben zu, die Lesen und Schreiben erforderten, beispielsweise die wichtige Aufgabe der Aufrechterhaltung eines adligen Korrespondenznetzwerkes. Elaborierte Fertigkeiten wie das Beherrschen der typisch barocken Formulierungsetikette, die Verwendung fremdsprachlicher Terminologie oder die fehlerfreie Rechtschreibung waren dafür aber nicht zwingend notwendig. Folglich geben die Schriftzeugnisse von adligen Frauen der Frühen Neuzeit einen unmittelbaren Einblick in die gesprochene Sprache der damaligen Zeit und Regionen und machen den hohen Normierungsgrad der geschriebenen Sprache ihrer männlichen Zeitgenossen dadurch gerade erst sichtbar.

Im Falle der Briefe Erdmuthe Benignas zeigen sich folgende **Schreibcharakteristiken**:

- durch Spatien abgetrennte Silben (insbesondere Präfixe) innerhalb der Zeile
- zahlreiche ‚fehlerhafte‘ Wörter, auch unterschiedlicher Schreibweisen gleicher Wörter auf gleicher Briefseite
- sehr willkürliche Interpunktion; Satzpunkte fehlen nahezu gänzlich; Kommata erscheinen (keineswegs regelmäßig) stellvertretend für Satzpunkte und oft stark abweichend von damaliger gelehrter als auch moderner Interpunktionspraxis
- heterogene Schreibweise gleicher Buchstaben, Vermischung von Kurrent und Antiqua in einem Wort

- kein Majuskelgebrauch zur Kennzeichnung von Satzanfängen, Feiertagen oder Gottesnamen; vereinzelt lediglich bei Anrede und Personennamen auftretend
- für die Geschäftsschrift ausgebildeter Zeitgenossen unüblich weitgezogene und runde Buchstaben
- chirographische Auffälligkeiten, die vermutlich durch die gebotene Eile und Arbeitsbelastung der Regentschaft einhergingen wie Streichungen, Überschreibungen, nachträgliche Ergänzungen und Marginalien

3. Projektumsetzung

3.1. Konzeption und Leitung

Mit der Konzeption und Leitung des Projektes wurde ein **wiss. Mitarbeiter** des Lehrstuhls betraut. Dazu gehörte eine zunächst intensive Auseinandersetzung mit Einsatzmöglichkeiten, Bedingungen und Arbeitsweisen von Transkribus. Auf deren Grundlage wurden das zu bearbeitende Quellenmaterial ausgewählt, vorverarbeitet und in Transkribus importiert, die Anforderungen der Mitarbeiterstellen formuliert und die erforderliche Projektlaufzeit kalkuliert. Während des Projektes kam dem wiss. Mitarbeiter insbesondere die Koordination und Betreuung der wiss. Hilfskräfte zu. Dies war insbesondere in der Frühphase des Projektes erforderlich, da zunächst beispielhafte Auszeichnungsvorlagen entwickelt, ein gemeinsamer Projektworkflow gefunden, unvorhergesehene Schwierigkeiten gelöst und editorische Entscheidungen getroffen werden mussten.

Zudem hatte sich der enge und sehr angenehme Kontakt zwischen dem wiss. Mitarbeiter und dem **Transkribusteam**, namentlich Tobias Hodel (Staatsarchiv des Kantons Zürich), als für das Projekt sehr wichtige und förderliche Aufgabe herauskristallisiert. Denn häufig ergaben sich Fragen erst während der detaillierten Bearbeitung des Quellenmaterials mit der Software, deren Dokumentation allerdings bereits den Großteil an Fragen beantwortet. Die adäquate Beantwortung des verbliebenen Restes sehr spezifischer Fragen ließ den regelmäßigen Kontakt mit einer Person bei Transkribus, die sowohl mit den technischen Gegebenheiten vertraut als auch den geisteswissenschaftlich-editorischen Herausforderungen solcher Projekte erfahren ist, sinnvoll erscheinen. Während der 2 ½ Monate wurden in zahlreichen Telefonaten und über einem Dutzend Emailwechseln sowohl grundlegende als auch sehr spezifische Fragen diskutiert und gelöst. Die Mails wurden durch Transkribus umgehend beantwortet und die Trainingsmodelle und das Wörterbuch schnellstmöglich trainiert (innerhalb weniger Tage), sodass keine verzögernden Unterbrechungen des Projekts entstanden.

3.2. Trainingsdatenerstellung (Ground-Truth)

Die Erstellung der Trainingsdaten oblag **zwei wiss. Hilfskräften**. Ihre im Rahmen der 2 ½ Monate durchgeführten Tätigkeiten umfassten:

- Vertraut-Machen mit und Austesten von Arbeitsweise und Funktionen der Software; die Vorteile vieler Funktionen werden erst in der gezielten Anwendung während des Projektes wirklich deutlich
- Segmentierung/Auszeichnung von Schriftbereich und Zeilen der Digitalisate
- Händische Nachbearbeitung der Zeilenerkennung; im Falle der CITlab-Advanced-Methode war dies allerdings nur vereinzelt notwendig; bei den älteren Layouterkennungsmethoden wiederum ging das vollständig händische Auszeichnen zuweilen schneller als die Korrektur der automatisch erkannten Linien
- zeilengenaue Übertragung der vorliegenden Transkriptionen in Transkribus; hierfür wurde mit den HTML-Vorlagen der Erdmuthé-Edition gearbeitet

- Kontrolle der Transkriptionen auf Korrektheit; dabei auch Überprüfung und ggf. Korrektur der Übernahme von Sonderzeichen
- Aussondierung nicht vorlagengetreuer Zeichen, die sich durch die Edition der Briefe ergaben (z.B. Fußnoten, eckige Klammern, im Fließtext platzierte Einfügungen von ursprünglich über, unter oder neben der Zeile stehender Wörter; Symbole für Personen, Orte und Institutionen)
- Sicherstellung der korrekten Verarbeitungsreihenfolge durch die Software (Hilfreiche Ansicht: „Show Lines Reading Order“ (Auge-Symbol)); Beim ‚händischen‘ Einfügen von Zeilen konnte eine fehlerhafte Reihenfolge entstehen
- Dokumentation der Nutzung der Software und dabei auftretender Probleme und Verbesserungsmöglichkeiten

Dieser Workflow war keineswegs so linear wie angegeben. Er erforderte vielmehr wiederholte Testdurchgänge und Nachbearbeitungen, die sich aus während des Projektes auftretenden quellenspezifischen Fragen und Entscheidungen ergaben. Diese ‚**Testphase**‘ sollte keineswegs unterschätzt und in Projektplanungen zu knapp kalkuliert werden, da sie die anschließende Dauer und Qualität der Auszeichnung und damit die finale Erkennungsrate der HTR beeinflusst. Sie war zwar zeitintensiv, rentierte sich aber in Form einer anschließend deutlich kürzeren Erstellungszeit des Trainingssets, als es ohne einen einmal gefundenen, stabilen und verbindlichen Workflow der Fall gewesen wäre. So wurde der überwiegende Teil des dadurch qualitativ hochwertigen und gleichmäßig ausgezeichneten Trainingssets erst in den letzten Wochen der Projektlaufzeit generiert.

3.3. Projektspezifische und -übergreifende Aspekte, Fragen und Entscheidungen

Während der Arbeit mit der Software an den konkreten Briefen traten Fragen verschiedener Art auf, die einer Entscheidung bedurften. Sie werden, da sie auch für andere Projekte von Interesse sein könnten, nachfolgend dokumentiert:

3.3.1. Automatische vs. händische Layoutanalyse

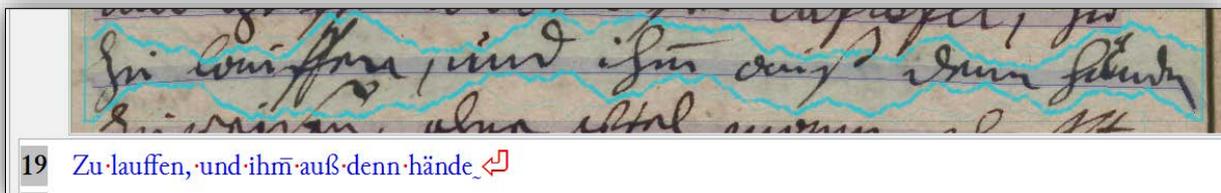
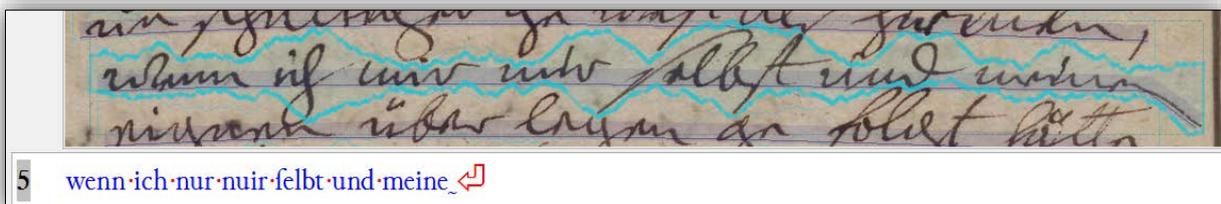
Die Layoutanalyse, sprich die Auszeichnung von Schriftbereichen und Textzeilen, ist ein entscheidender Vorverarbeitungsschritt, dessen Ausführung die zu erwartende Zeichenerkennungsrate maßgeblich beeinflusst. Transkribus stellt dafür eine automatisierte und eine händische Bearbeitungsoption bereit. Welche davon angewendet wird, spielt in technischer Hinsicht und mit Blick auf die Erkennungsrate zunächst keine Rolle. Entscheidend ist ein sauberes Ergebnis der gewählten Methodik. Besonders gute Ergebnisse lieferte die automatisierte Erkennungsmethode „**CITlab Advanced**“. Sie erforderte im Vergleich mit den beiden älteren Layoutanalyse-Tools die wenigsten Nachbearbeitungen. Ggf. notwendige händische Nachkorrekturen, gerade was vorzeitige Zeilenumbrüche, Hinzufügungen und Marginalien betrifft, führten zuweilen zu Fehlern bei der Verarbeitungsreihenfolge. Das wird sichtbar, wenn die Zeile in der Listenansicht nicht dort eingefügt wird, wo man sie auf dem Faksimile zuvor platzierte, was wiederum schnell zu Fehlern bei der händischen Zuordnung von Transkriptions- und Faksimilezeile führen kann. Entsprechend empfiehlt es sich, die Verarbeitungsreihenfolge bei Nachkorrekturen stets zu überprüfen. Eine dafür sehr nützliche Ansicht, die über das Faksimile gelegt werden kann und damit den Blick in die Listenansicht erspart, bietet die Ansichtsoption „Show Lines Reading Order“ (siehe Augesymbol).

3.3.2. Buchstabensuspension am Zeilenende

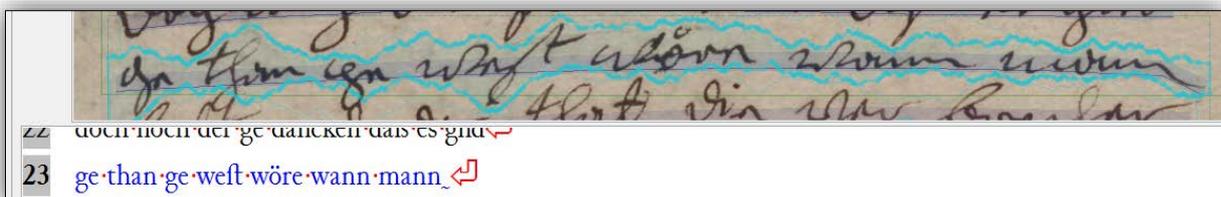
Handgeschriebene Briefe der Frühen Neuzeit weisen häufig auslaufende Linien statt graphisch sauber ausgeführte Buchstaben am Zeilenende auf. Dies betrifft meist einzelne oder mehrere Buchstaben wie „n“ oder „en“, kann sich aber auch auf längere Buchstabenketten und Silben erstrecken. Dies macht jeden Zeilenausläufer zu einem aus dem Satzkontext zu interpretierenden Einzelphänomen, das nicht einfach automatisch aufgelöst werden kann, da sich das Graphem selbst zwar kaum verändert, die

dahinter stehende Semantik hingegen schon. Im Falle zeichengenauer Transkription sollte dieser Interpretationsprozess zudem zusätzlich zur Dokumentation des diplomatischen Textbefundes, nicht aber diesen ersetzend stattfinden. Die Dokumentation kann, wie im vorliegenden Beispiel, in Form eines Platzhalters erfolgen, idealerweise einem graphisch möglichst nahe an das ursprüngliche Zeichen herankommenden Unicode-Zeichen, wie bspw. der tiefstehenden Tilde (˷). Aus diesen editorischen Gründen einerseits und technischen Bedenken andererseits, die graphisch gleichbleibende Suspension könnte bei sich ändernder Semantik eine mögliche Fehlerquelle der HTR darstellen, wurden diese nicht etwa aufgelöst, sondern als Tilde transkribiert. Rein technisch betrachtet sind jedoch letztlich beide Methoden – Tilde und stillschweigende Auflösung – fehleranfällig, sodass das Ergebnis ähnlich ausfallen dürfte. In beiden Fällen kann die HTR Stellen zwar ähnlicher Erscheinung jedoch unabhängig eines tatsächlichen Zeilenausläufers als der dafür hinterlegte Platzhalter bzw. eine hinterlegte Auflösung transkribieren. Empfehlenswert ist daher vielmehr eine **pragmatische Lösung**, die in den Workflow des Projektes passt (sprich der gewünschten Ausgabe und Vorarbeiten entspricht), möglichst unproblematisch einzugeben ist und bei der Korrektur erkennbar bleibt. Daher wurde die Platzhalter-Variante angewendet.

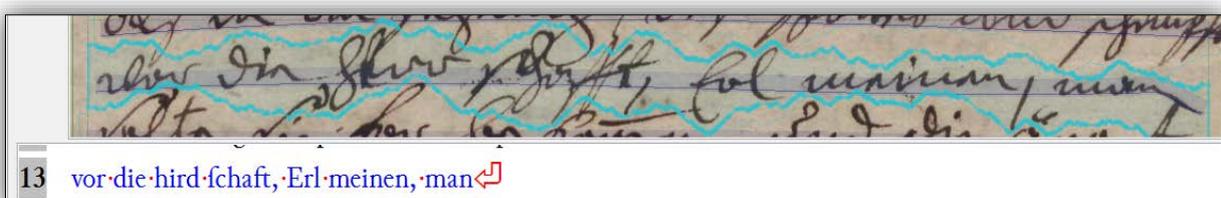
Im Ergebnis (ohne Wörterbuchabgleich) erkennt die HTR die Abbruchlinien in der Regel erstaunlich zuverlässig und transkribiert diese in Form der gewünschten tiefstehenden Tilden:



Zuweilen ‚schießt‘ die HTR aber auch über das Ziel hinaus ...



... oder erkennt die auslaufende Linie nicht:

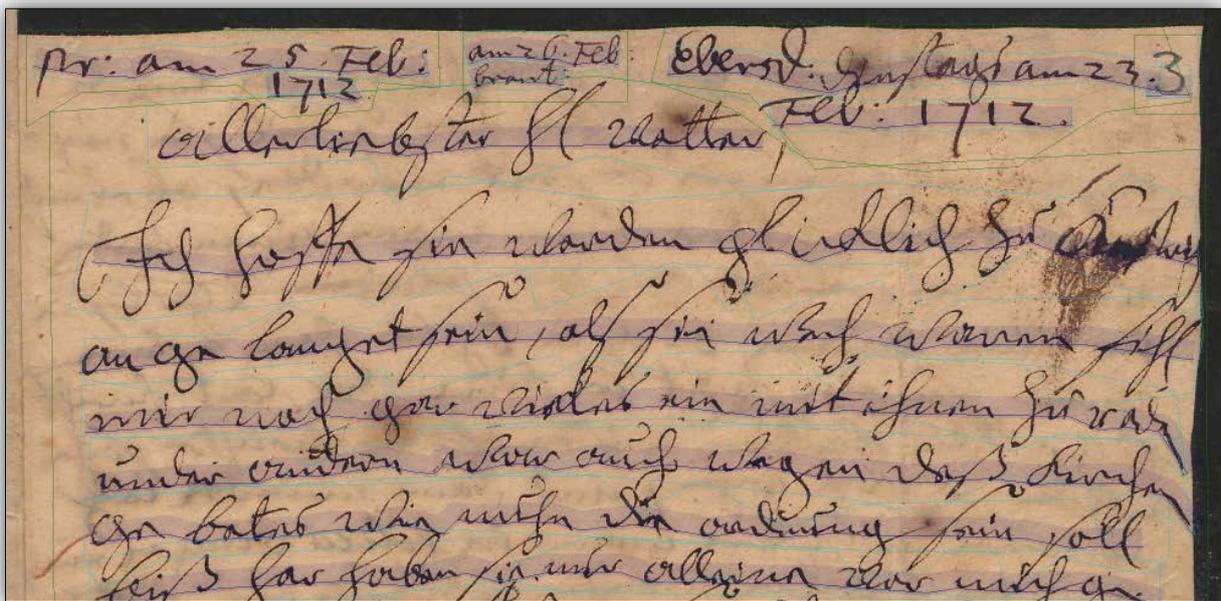


3.3.3. Zeilenwechsel vor Blattende

Ein häufig vor allem zu Beginn oder Ende der Briefe auftretendes Phänomen betrifft den vorzeitigen Zeilenwechsel noch vor dem Blattende und das Anführen eines neuen Gedankens auf gleicher Höhe, was einer Art Spaltenschreibweise ähnelt. Verschiedene Möglichkeiten des Umgangs damit standen zur Diskussion:

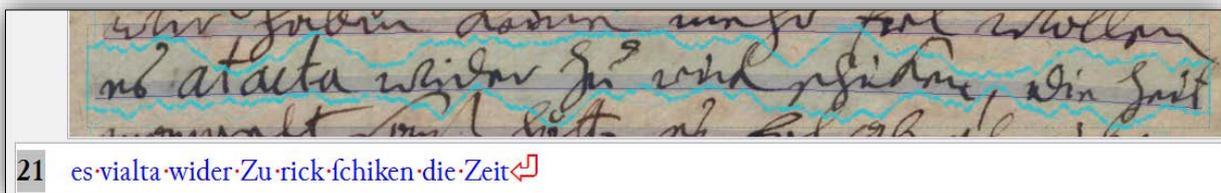
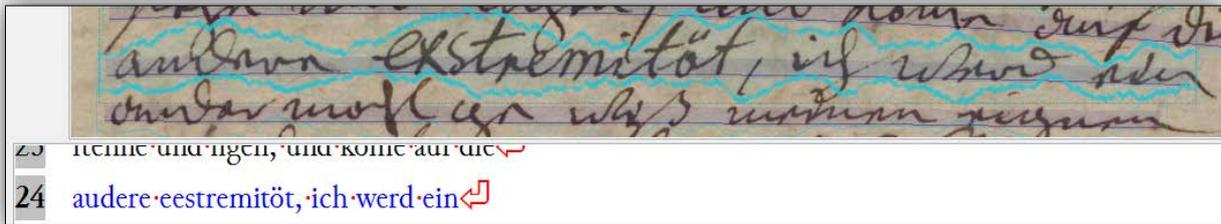
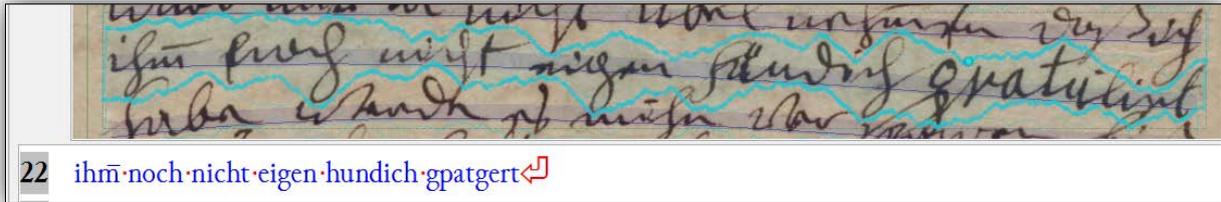
1. Jede ‚Spalte‘ als separate Textregion auszuweisen und die Zeilenauszeichnung entsprechend des syntaktisch-semanticen Textbefundes vorzunehmen.
2. Eine spaltenübergreifende Textregion zu definieren und die Zeilenwechsel ungeachtet des inhaltlichen Textbefundes erst am Blattende vorzunehmen.
3. Auszeichnung als Tabellenspalte statt Textregion und Zeilenauszeichnung entsprechend 1.

Letztendlich wurde sich für Variante 1 entschieden, da die Ausgabe als Tabelle bzw. als durchgehende Zeile im Anschluss schwieriger weiterzuverarbeiten gewesen wäre. Dabei konnten sich die verschiedenen Textregionen auch durchaus überschneiden:



3.3.4. Unterscheidung kurrenter und lateinischer Buchstaben

Insbesondere in barocken Schriften war ein üblicher Usus der Zeit, fremdsprachliche Begriffe in überwiegend neugotischen/kurrenten Texten als Antiqua-Buchstaben wiederzugeben. Es stellte sich daher die Frage, ob die explizite Auszeichnung als Antiqua-Buchstaben ein empfehlenswerter Mehraufwand für eine möglichst hohe Erkennungsrate sei. Da die Unterscheidung zwischen lateinischen und kurrenten Buchstaben – je nach Größe des Trainingssets – in anderen Transkribusprojekten bereits recht gut funktioniert, wurde sich gegen ein zusätzliches Taggen dieser Begriffe entschieden, zumal die HTR das Tagging bei der Erkennung – zumindest derzeit – auch gar nicht einbezieht. Im hiesigen Projekt ergab die anschließende Qualitätskontrolle allerdings deutliche **Schwächen** bei der Erkennung:



Der Grund dafür ist, dass die Anzahl an Antiquabuchstaben in den Briefen Erdmuthes noch zu gering ist, als dass genügend Trainingsmaterial für eine korrekte Erkennung zur Verfügung stünde. Eine explizite Auszeichnung als Antiqua hätte also in keinem Fall ein besseres Ergebnis produziert. Auch wenn sich dies in der Zukunft durch technische Modifikationen eventuell einmal ändern sollte, so ist zudem stets der Kosten-Nutzen-Aufwand angesichts der – je nach Text und Schreiber/in – meist nur geringen Häufigkeit von Antiquabuchstaben zu berücksichtigen.

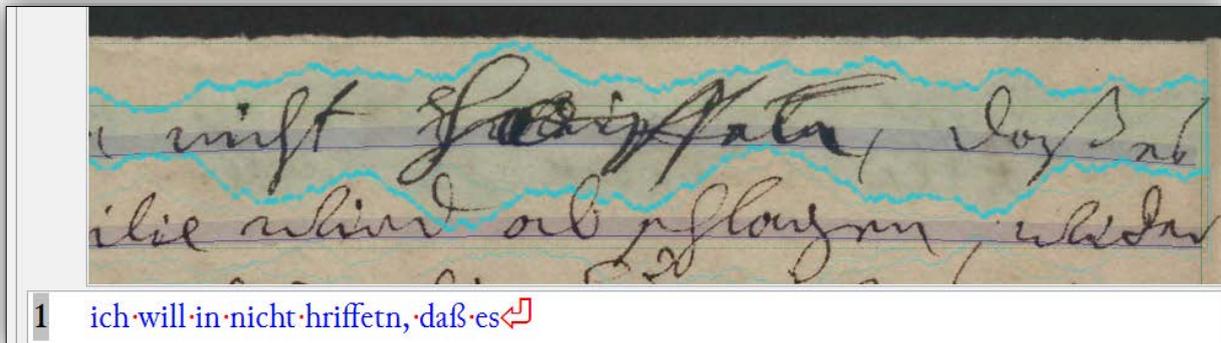
3.3.5. Überschriebene Wörter und Buchstaben

Die Briefe Erdmuthes Benignas lassen die oft gebotene Eile der Schreiberin unter anderem an den vorgenommenen Streichungen und Ergänzungen erkennen. Einer dieser Korrekturtypen stellen unmittelbar auf einem anderen Buchstaben oder Wort vorgenommene Überschreibungen dar. Die HTR kann – im Gegensatz zur/m menschlichen Nutzer/in – verständlicherweise von sich aus keine Unterscheidung zwischen überschriebenem und übergeschriebenem Buchstaben treffen, sondern versucht beides als ein Buchstabe zu lesen. Folglich musste eine Entscheidung getroffen werden, wie mit den betroffenen Stellen bei Layoutanalyse und Transkription umgegangen wird. Hier boten sich zwei Möglichkeiten an:

1. Auslassen der Layoutanalyse an betreffender Stelle mit zwangsläufiger Unterbrechung vor und Neuansetzung der Baseline nach dieser Stelle, wodurch allerdings zwei Linien gebildet werden.
2. Auszeichnung mit durchgehender Baseline und Transkription des jeweils letzten Bearbeitungsstandes, sprich der übergeschriebenen Zeichen als die mutmaßlich am besten zu erkennenden Zeichen.

Aus Zeiteffizienz-Gründen und um auch hier Grenzen der HTR zu testen und dabei auf möglichst wenig Textmaterial zu verzichten, wurde Option 2 gewählt. Diese ignoriert das Überschriebene, auch wenn dies deutlicher sichtbar war als das Übergeschriebene. Entsprechend hat das Erdmuthes-Modell im Test-Material erwartungsgemäß Schwierigkeiten betroffene Stellen zu erkennen.

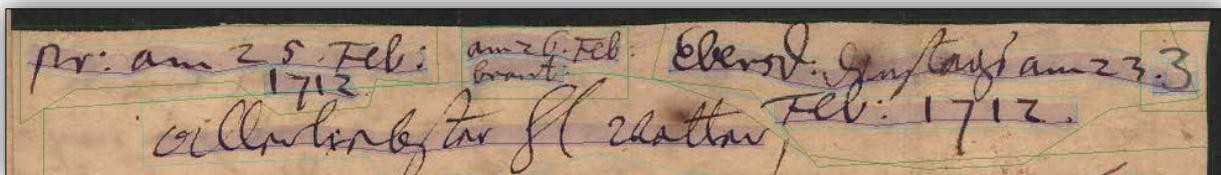
Zusätzlich könnten die transkribierten Zeichen noch mit „unclear“ getagged werden. Damit wird die Zeile für das Training ausgelassen. Dies eignet sich, wie Option 1 auch, vor allem dann, wenn ohne die Auslassungen noch genug Trainingsmaterial verfügbar ist.



Überschrieben: „zweifeln“; übergeschrieben: „hoffen“; erkannt: „hriffetn“.

3.3.6. Haupthand und andere Schreibhände

Der überwiegende Teil der Schrift innerhalb der Briefe entstammt der Hand Erdmuthe Benignas, deren Handschrifterkennung das Ziel des Projektes war. Neben dieser Haupthand treten aber auch Handschriften weiterer Personen auf, insbesondere die des Briefempfängers in Form von Eingangsvermerken oder schlagwortartigen Inhaltszusammenfassungen. Streng genommen führt die Transkription fremder Hände zu einer ‚**Verunreinigung**‘ und damit zu potentiellen Fehlerquellen des Trainingssets, von dem meist lediglich die Haupthand von Interesse ist. Auch liefern fremde Hände, die nur gelegentlich auftreten, **zu wenig Trainingsmaterial**, als dass sie zuverlässig erkannt werden könnten. Letztlich ist es also eine Frage, wie umfangreich die Ergänzungen fremder Hände sind. Im vorliegenden Projekt war der Einfluss von Fremdhänden so gering, dass das Ergebnis dadurch nur im minimalen und daher vernachlässigbaren Prozentbereich beeinflusst wird. Auch die Erfahrung von Transkribus mit anderen Projekten zeigt, dass bei der Kombination von Modellen mit zwei Händen die Erkennung der größeren (Ground-Truth-)Hand durch die Kombination nicht bzw. höchstens minimal (im 0,5%-Bereich) verschlechtert wird. Erst ab ca. fünf Händen ist teilweise eine Verschlechterung spürbar.



Das Beispiel zeigt die oft undurchsichtige Verquickung zweier Hände. Während Ort, Wochentag und Anrede von der Schreiberin stammen, wurden Präsentations- und Antwortvermerk sowie die Datumsergänzung von dem Adressaten hinzugefügt.

3.3.7. Einbezug von Wörterbüchern

Wie aus OCR-Verfahren bekannt, bietet auch Transkribus die Möglichkeit, Wörterbücher zur HTR-Erkennung der Wörter hinzuzuziehen und so die Erkennungsrate weiter zu erhöhen. Dadurch wird nicht länger nur die wahrscheinlichste Zeichenerkennung ausgegeben, wie das ohne Wörterbucheinsatz der Fall ist, sondern das wahrscheinlich erkannte Wort zusätzlich auch mit dessen Vorhandensein in einem ausgewählten Wörterbuch abgeglichen. Die noch nicht normierte Orthographie frühneuzeitlicher Briefe im Allgemeinen und die eigentümliche Schreibweise Erdmuthe Benignas im Besonderen (s.o.)

stellen dieses Prinzip allerdings vor das Problem des Fehlens eines optimalen Wörterbuchs. Daher kann entweder auf ein bereits vorhandenes Wörterbuch eines anderen Projekts zu einer idealerweise ähnlichen Zeitepoche mit ähnlichem Schreibduktus zurückgegriffen werden. Oder es wird ein ‚eigenes‘, **korpuspezifisches Wörterbuch** generiert, das auf der von dem/der Schreiber/in selbst verwendeten Rechtschreibung basiert und möglichst viele typische Schreibvarianten eines Wortes umfasst, idealerweise ohne große Variabilität der einzelnen Lemmata. Mit zunehmender Größe des Wörterbuches geht nämlich eine sukzessive Verlangsamung der Erkennung einher. Ein Wörterbuch, das aus 150.000 Seiten erstellt wurde, benötigt beispielsweise eine 60 Minuten (!) längere Verarbeitungsdauer und liefert eine CER-Verbesserung von nur etwa 1%.

Für die Generierung eines Wörterbuches muss das Transkribus-Team kontaktiert werden, das dieses in kurzer Zeit erstellt und bereitstellt. Danach ist es neben allen bisher generierten Wörterbüchern zu finden unter „Text Recognition Configuration“. Im Falle der Erdmuthebriefe waren die Eigenheiten der Schreibung zu disparat, als dass der Einbezug eines bereits erstellten Wörterbuches (z.B. Wörterbuch der Senckenberg-Handschriften) ein besseres Ergebnis gebracht hätte. Fehlerkennungen verursachte das Senckenberg-Wörterbuch unter anderem aufgrund der untypischen Silbenabtrennung und der Geminationsstriche der Erdmuthe-Briefe. Das Anlegen eines projekteigenen Wörterbuchs erschien daher erfolgversprechend. Und in der Tat führte der Einbezug eines Wörterbuchs auf Grundlage der Erdmuthe-Briefe (**Erdmuthe_M2.dict**) noch einmal zu einer weiteren Verbesserung der Erkennungsrate auf einer Testseite (s.u. „Endergebnis“) bei zugleich nicht signifikant erhöhter Verarbeitungsdauer. Dies war angesichts der spezifischen Silbentrennung Erdmuthes einerseits zu erwarten, aufgrund der in sich sehr wechselhaften Schreibweise auch gleicher Wörter andererseits aber auch etwas überraschend.

3.3.8. Offline-Arbeiten

Transkribus ermöglicht durch die Client-Server-Struktur die zeit- und ortsungebundene Bearbeitung der Dokumente. Auch **kurzzeitige Unterbrechungen** der Internetverbindung sind kein Problem. Transkribus erstellt während dieser Phasen lokale Sicherungskopien, die bei Rückkehr besserer Internetverbindung und nach Bejahung einer automatischen Nachfrage verwendet werden können, um den zwischenzeitlichen Arbeitsfortschritt nicht zu verlieren. Allerdings kann Transkribus lediglich die aktuell angezeigte Seite beantworten. Soll eine neue Seite geladen werden, bedarf es des erneuten Internetzuganges.

Um das Arbeiten auch während **längerer Offlinephasen** zu ermöglichen, hatte es sich bewährt, die zu bearbeitenden Seiten in kleinere Pakete („Dokumente“) zu ‚schnüren‘ und hochzuladen. So lässt sich ein Teil der Dokumente projektintern für das Offline-Arbeiten, ein anderer für das Online-Arbeiten definieren. Dies wurde erforderlich, da innerhalb von Transkribus (offline) bearbeitete Seiten zwar im PAGE-MXL-Format lokal gespeichert und exportiert werden können. Beim nachträglichen ‚händischen‘ Server-Upload dieser Seiten werden allerdings Duplikate erstellt. Die dadurch einhergehende Versionsvielfalt unterschiedlicher Bearbeitungsstufen derselben Seite kann zu Verwirrungen und Missverständnissen führen, insbesondere bei kollaborativem Arbeiten.

3.3.9. Abstand der Baseline zum Zeichen

Ein nicht zu vernachlässigender Aspekt betrifft die **optimale Positionierung der Baselines**, um die so wichtige korrekte Zeilenerkennung und damit das Gesamtergebnis nicht zu gefährden. Vermieden werden sollten zu tiefe Baselines, sprich Leerräume zwischen Buchstaben und Grundlinie. Unproblematischer sind etwas höhere Baselines. Diese können zwar oberhalb der Grundlinien stehen und damit die Wörter durchlaufen. Sie sollten die Mittellänge allerdings nicht tangieren. Eine Möglichkeit, die anschließende korrekte Zeilenerkennung durch die HTR zu testen, besteht in der Anwendung eines beliebigen vorhandenen HTR-Modells. Bei einer optimalen Erkennung sollten die Zeilen vollständig und korrekt von **Polygonen** umschlossen sein („item visibility“ (Augesymbol) --> „show lines“).

3.3.10. Automatisierte Zuordnung von Text- und Bildzeilen

Neben der händischen Zeilenzuordnung von Transkription und Digitalisat besteht die Möglichkeit der automatisierten Zuweisung mittels eines **Text-to-Image-Tools (t2i)**. Das Tool wurde durch das CITlab Team, insbesondere Gundram Leifert, der Universität Rostock entwickelt. Das Team zeichnet sowohl für die aktuell laufende HTR sowie die Layouterkennung verantwortlich. T2i setzt beide Technologien kombiniert ein, indem ein Bild zuerst hinsichtlich des Layouts erkannt wird. Danach erfolgt die Erkennung mit einem bereits vorhandenen Modell (auch schlecht funktionierende Modelle können eingesetzt werden). Die Lesung des Modells wird jedoch nicht ausgegeben, sondern mit Textdateien abgeglichen, die den zu erkennenden Text enthalten. (Varianten-)Lesungen des Modells die mit Text aus der Textdatei übereinstimmen, werden anschließend zugewiesen. Der Anwendung von t2i sind aktuell nur beschränkte Ressourcen (Rechnerkerne) zugeteilt, weshalb die Zuordnung andauert und nicht für alle freigegeben ist. Mit einem halbwegs funktionierendem Modell (Fehlerrate bis zu 40% CER) werden ca. 95% der Zeilen korrekt erkannt. Das Ausgangsformat der Transkriptionen ist dabei weniger relevant und kann sowohl .xml, .docx als auch .txt umfassen, wenngleich Letzteres das bevorzugte Datenformat ist. Digitalisat- und Transkriptionsseite sollten zudem bereits einen identischen Namen tragen. Im Falle des hiesigen Projektes hätten die bereits vorliegenden Transkriptionen allerdings noch bereinigt werden müssen (Fußnoten, Icons, eckige Klammern und weitere nicht vorlagengetreue Zeichen und Anordnungen). Auf die Anwendung des Tools wurde daher verzichtet, sodass die Qualität der Ergebnisse von t2i nicht geprüft werden konnte.

4. Projektergebnisse und deren Weiterverwendung

In technischer Hinsicht ist die **Character Error Rate (CER)** ein hilfreiches Maß, mit dem das Ergebnis eines trainierten Erkennungsmodells eines Transkribusprojektes beurteilt und verglichen werden kann. Sie gibt an, wieviel Zeichen die HTR auf Grundlage eines Trainingssets fehlerhaft erkannt hat. Die Definition von Zeichen ist für CER streng, es fallen darunter auch Leer- und Satzzeichen. Zum Vergleich: Die OCR-Erkennung moderner Drucke erreicht einen CER von unter 0,5%. Für Anwendungen der HTR auf Handschriften ist von Bestwerten um 3% CER (3 Fehler pro 100 Zeichen) auszugehen.

Zwischenergebnis

Um das zu erwartende CER-Endergebnis bereits vor Projektende einschätzen zu können, wurde nach der Bearbeitung von 51 Seiten (8.839 Wörter) ein erstes Trainingsmodell (Erdmuthe_M1) erstellt. Das Trainieren eines solchen Modells ist innerhalb von drei Tagen möglich und wird vollständig durch das Transkribusteam vorgenommen. Nach dem Trainieren steht das HTR-Modell unter „Tools“ innerhalb der Software zur Verfügung. Die Anwendung von Erdmuthe_M1 auf eine vorher unbearbeitete Testseite ergab eine CER von **10,62%**, sodass – statistisch betrachtet – fast 9 von 10 Zeichen bereits richtig erkannt wurden. Im Trainingsset war die Fehlerrate mit 3,01% noch einmal deutlich geringer, was darauf hindeutete, dass mit mehr Trainingsmaterial noch bessere Ergebnisse erzielt werden können. Dies war ein für das Projekt sehr hilfreicher Zwischenstand, der einen optimistischen Blick auf das Endergebnis freigab.

Endergebnis

Am Ende der Projektlaufzeit wurde das finale Trainingsset erstellt (Erdmuthe_M2). Es basiert auf 290 Seiten (48.277 Wörter). Das Testset wurde nun auf vier Seiten vergrößert und wies eine CER von **8,89%** auf. Damit konnte im Vergleich zum ersten Modell noch einmal eine Verbesserung von fast 2% erreicht werden. 91 von 100 Zeichen wurden damit im Durchschnitt durch die HTR korrekt automatisch transkribiert.

Der Vergleich zwischen den Modellen Erdmuthe_M1 und Erdmuthe_M2 verdeutlicht auch, dass die Größe von bearbeitetem Trainingsset und Erkennungsrate sich nicht proportional verhalten, sondern

Erfahrungs- und Ergebnisbericht zum Transkribus-Projekt

vielmehr exponentiell. Während bereits mit recht wenigen Trainingsdaten gute Ergebnisse unterhalb einer Fehlerquote von 10% erzielt werden können, erfordert das Verringern der letzten übrigen CER-Prozentpunkte ein ungleich größeres Trainingsset.

Dass es sich bei den genannten CER-Raten um Mittelwerte handelt, die sich aus recht großen Spannen sowohl besser als auch schlechter erkannter Seiten ergeben, soll anhand des nachfolgenden Beispiels verdeutlicht werden:

Ergebnis Modell „Erdmuthe_M2“ (ohne Wörterbuch)	Ergebnis Modell „Erdmuthe_M2“ (mit Wörterbuch)	Korrekte ‚händische‘ Transkription
<p>bekitt an 13 Mani aller liebfter H vetter 171 EJchbe ruffe mich widernn auf Carln der wird von allen Erl be richt er ftattet haben, will dur wegen deß groffen tetalse Zu wetzler dißes fagen, daß fo gern allich dife unöhtige en kosten er baret hätte, fo hatt es doch nicht wohl anders fein können, in dem es offendlich draußien in bey fein viller ab gelesen wird, und ohne dem im reich sie uißer hauß gar gering halte weil sie fo gar keine nach richt vom reiffchen hauß haben, und es Zu erft an die familie Zu berechten, ver hinderde die ohne dem fo lang ver Zogne lehn fuchnug, hätten wir es eine weille auf folche ard mit denn herren titil auß ferdigen laßen dirfte sich din familie über unß be schworen, daß wir Zu er sparung der kosten der gleichen precatie vier ur fachtet hätten,</p>	<p>er bereit an 13 mann aller liebfter H vetter 171 EJchbe ruffen mich weidern auf Careln der wird von allen Erl be richt er ftattet haben, will nur wegen deß groffen titoll Zu wetzler dißes fagen, daß fo gerne allich dife unöhtige ein kosten er sparet hätte, fo hatt es doch nicht wohl anders fein können, in dem es hoffendlich draußen in bey fein viller ab gelesen wird, und ohne dem im reich sie unßer hauß gar gering halten weil sie fo gar keine nach richt vom reiffchen hauß haben, und es Zu erft an die familie Zu berichten, ver hinderde die ohne dem fo lang ver Zognen lehn fuchung, hätten wir es eine weille auf folche ard mit denn herren titil auf ferdigen laßen dirfte sich die familie über unß be schworen, daß wir Zu er sprung der kosten der gleichen precatie ver urfachtet hätten,</p>	<p>227 beant am 13. Maii aller liebfter H vetter 1712. Jch be ruffe mich widerum auf Careln der wird von allen Erl be richt er ftattet haben, will nur wegen deß groffen tituls Zu wetzler dißes fagen, daß fo gerne allich dife unöhtige un kosten er sparet hätte, fo hatt es doch nicht wohl anders fein können, in dem es offendlich draußen in bey fein viller ab gelesen wird, und ohne dem im reich sie unßer hauß gar gering halte weil sie fo gar keine nach richt vom reiffchen hauß haben, und es Zu erft an die familie Zu berichten, ver hinderde die ohne dem fo lang ver Zogne lehn fuchung, hätten wir es eine weille auf folche ard mit denn herren titil auf ferdigen laßen dirfte sich die familie über unß be schwören, daß wir Zu er sparung der kosten der gleichen preiutic ver ur fachtet hätten,</p>

Diese Testseite zeigt mit einer CER von 6,03% ohne Wörterbuch (von insgesamt 646 Zeichen (ohne Leerzeichen) wurden 39 fehlerhaft erkannt, s. rote Zeichen) eine deutlich niedrigere Fehlerrate als der Gesamtdurchschnitts-CER-Wert von 8,89%. Mit Wörterbuch (Erdmuthe_M2.dict) reduzierten sich die Fehler noch einmal um neun Stück und damit auf eine Fehlerrate von **4,64%** auf der Testseite. Ein Grund für das deutlich bessere Ergebnis im Vergleich mit dem Gesamtdurchschnittswert des Projektes sind unter anderem die relative Gleichmäßigkeit und Sauberkeit von Schriftbild und Schriftträger dieser Testseite. Für dieses Beispiel wurde die automatische Layoutanalyse (CITlab Advanced) zwar zuvor minimal händisch nachbearbeitet. Doch auch ohne Nachkorrektur entstanden lediglich zwei Fehler mehr.

Auffällig ist zudem, dass die Fehler in gehäufte Form bei fremder Schreiberhand und Antiquaschreibweise, für die dem Modell zu wenig Trainingsmaterial zur Verfügung stand, sowie in Folge von Durchstreichungen auftraten. Die von Erdmuthe sauber geführte Hand wies hingegen nur etwa ein bis zwei

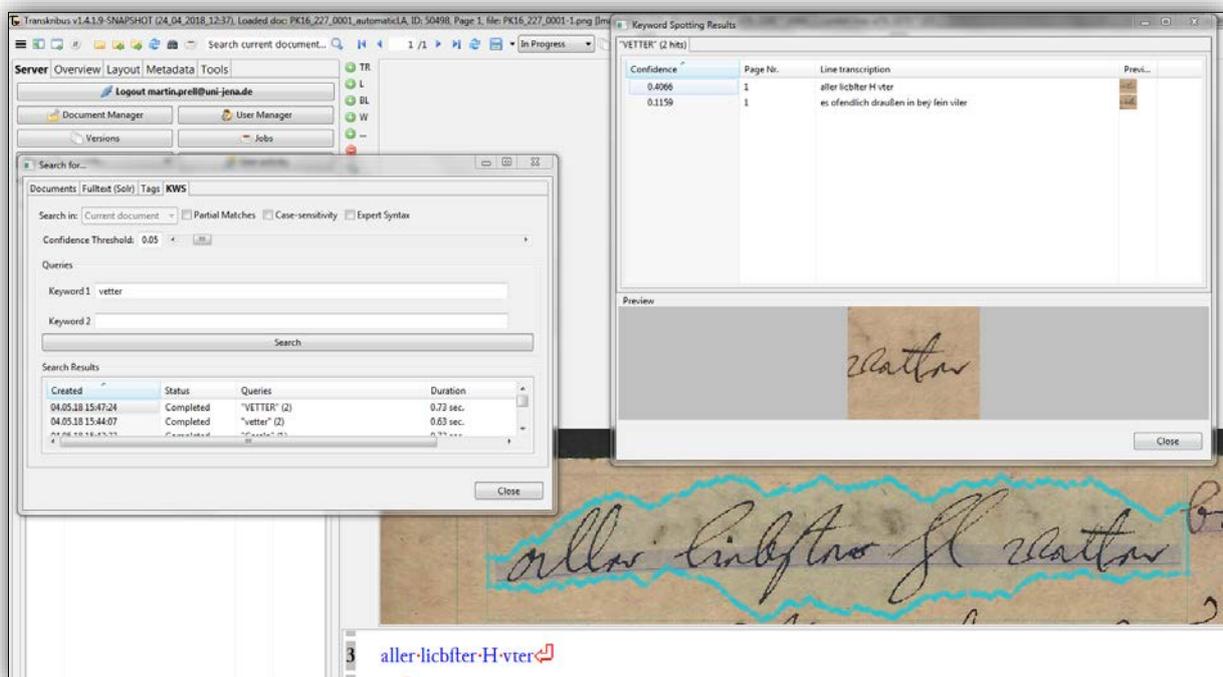
Zeichenfehler pro Zeile auf. Das Beispiel verdeutlicht damit, dass bei einem ausreichend großen Trainingsset weniger die Handschrift an sich Erkennungsschwierigkeiten verursachte, sondern vielmehr die durch die Bearbeitungsfreiheit des Mediums einhergehenden Sonderfälle.

Deutlich wurde zudem, wie wichtig die korrekte Erkennung der Polygone und damit die richtige Einzeichnung der Baselines (s.o.) ist. Wo die Polygone nicht den gesamten Buchstaben umfassten (bspw. bei u-Überstrich) wurden die Buchstaben nicht korrekt erkannt.

Hingewiesen sei zudem auf eine neue HTR-Version (HTR+), die im Sommer von Transkribus bereitgestellt wird und noch einmal eine deutliche Verbesserung der Erkennung ermöglichen dürfte.

Ergebnisverwendung

Transkribus bietet seit kurzem die Möglichkeit des **Keyword-Spotting** in Dokumenten. Diese bezieht nicht nur die von der HTR final ausgegebene, wahrscheinlichste Erkennung eines Zeichens ein, sondern auch die weiteren von der HTR als weniger wahrscheinlich ermittelten Lesungen. Damit handelt es sich bei der KWS um eine äußerst nützliche da fehlertolerante Suche, die dem Nutzer die Möglichkeit gibt, mit den maschinellen Ergebnissen noch gezielter arbeiten zu können. So kann trotz zunächst vermeintlich falscher Erkennung eines Wortes das korrekte Wort dennoch gefunden werden. Bei Suche eines Keywords werden alle Treffer absteigend sortiert nach dem Grad an Gewissheit/Verlässlichkeit der Erkennung (Wert zwischen 0 und 1) aufgeführt und die identifizierten Textstellen mittels Thumbnail übersichtlich präsentiert. Der Grenzwert an Gewissheit bis zu dem Transkribus alle Treffer eines Suchwortes ausgeben soll, kann durch die Nutzenden selbst festgelegt werden. Der Einbezug regulärer Ausdrücke ist ebenfalls möglich, genauso wie die Option, auch nur teilweise Übereinstimmungen von gesuchtem und gefundenem Wort anzuzeigen.



Beispiel für die Findbarkeit vermeintlich falsch erkannter Wörter mittels KWS. Statt „veter“ wurde durch die HTR zwar „vter“ als die wahrscheinlichste Lesung transkribiert. Die KWS hingegen macht sichtbar, dass sich hinter „vter“ mit einer Gewissheit von 40% auch der Begriff „vetter“ verbergen könnte.

5. Subjektive Einschätzungen und Ausblicke

Das hier beschriebene Projekt verdeutlicht, dass Transkribus auch in Anwendung auf kleinere Briefkonvolute und eine sowohl für den menschlichen Leser als auch die maschinelle Erkennung herausfordernde Schrift bereits sehr gute Ergebnisse produziert, die sich für die **inhaltliche Erfassung** und das **gezielte Durchsuchen** auf bestimmte Begriffe hin eignen. Ohne Nachbearbeitung liefert Transkribus noch keine an durch Fachwissenschaftler/innen oder externes Double-Keying erstellte heranreichende Transkriptionen. Für viele Handschriften wird die annähernd fehlerfreie rein automatische Erkennung wohl auch niemals Realität, ist die Heterogenität und Qualität von Handschriften und Schriftträgern doch theoretisch genauso unbegrenzt wie der menschliche Geist an sich. Die/Den interpretierende/n Editor/in wird die Maschine daher keinesfalls ersetzen und damit weder ‚Heilsversprechen‘ noch Bedrohungsszenarien der Digitalisierung zumindest in dieser Hinsicht einlösen. Allein der Mensch kann, sofern er ebenfalls hervorragend ausgebildet ist, die letzten Prozentpunkte der CER tilgen. Dass aber – langfristig betrachtet – die nahezu fehlerfreie Texterfassung eines Großteils des in Gedächtniseinrichtungen befindlichen und häufig gänzlich unbekanntem textuellen Kulturerbes ohne aufwändige Nachbearbeitung in greifbare Nähe rückt, zeigen die erfolgreiche Entwicklungsgeschichte der Software und deren breite Anwendung in bereits zahlreichen Projekten. Transkribus muss folglich auch zukünftig stetig weiterentwickelt und am historischen Quellenmaterial getestet werden.

Angesichts einer sowohl im wissenschaftlichen als auch gesamtgesellschaftlichen Bereich wahrnehmbaren zurückgehenden Fähigkeit historische Handschriften zu lesen, beinhaltet diese Entwicklung das Potential zu einem **neuen Interesse an den Beständen von Archiven, Bibliotheken, Museen und Vereinen**, die sonst unbekannt und unbeachtet bleiben würden. Durch Projekte wie Transkribus dürfte sich also gerade nicht die zuweilen skizzierte Befürchtung der Verdrängung von traditionellen Aufgaben und Bedeutung von Kulturgutwahrenden Institutionen und den Stellen der Mitarbeiter/innen einstellen, sondern das Gegenteil. Dafür müssen die Gedächtniseinrichtungen aber auch gezielt aktiv werden und die inhärenten Chancen solcher Entwicklungen bewusst nutzen und fördern. Denn die langfristige Bewahrung von Kulturgut erfolgt nicht etwa über dessen Verschließen, Verbergen und Vergessen, sondern zuallererst über das Wecken von Interesse an diesem. Um erfolgreich und langfristig bewahren zu können, muss bekannt werden, was überhaupt bewahrt wird. Dafür ist die Inhaltserschließung essenziell. Transkribus bietet diese Möglichkeit in bereits sehr zuverlässiger Qualität. Durch die automatisierte Volltexterfassung verleiht es insbesondere gänzlich unerschlossenen oder gar unbekanntem Dokumenten kulturellen und letztlich sogar ökonomischen Wert.

Zwar mag die Anwendung von Transkribus auf serielle Quellen einen noch deutlicheren Mehrwert versprechen, stehen Briefe doch ohnehin häufig im wissenschaftlichen und öffentlichen Interesse und werden ediert. Doch durch ihre – im Vergleich mit Druckschriften – enorme Spanne an Gestaltungsfreiheit regen sie zu einer steten Verschiebung der Grenzen des technisch und editorisch Möglichen an.

6. Übersicht Projektdaten

Laufzeit: 15. Oktober bis 31. Dezember 2017

Personal: zwei wiss. Hilfskräfte (1x Bachelor, 1x Master) im Umfang von 40h/Monat und ein wiss. Mitarbeiter im Umfang von ca. 20h/Monat

Trainingsdaten: 66 Briefe Erdmuthe Benignas an Heinrich XXIV. Reuß-Köstritz = 290 Seiten, 48.277 Wörter

Durchschnittliche CER (Testmaterial, 4 beliebige Seiten): 8,89%

CER auf gezielt ausgewählter Testseite: 4,64%

Erfahrungs- und Ergebnisbericht zum Transkribus-Projekt

Ausstattung: Digitalisate und Transkriptionen der Briefe, Laptop (auch ältere Modelle einsetzbar), Transkribussoftware samt Transkribusaccount

Einer möglichst kurzen Projektlaufzeit, geringen Projektkosten und einem guten Projektergebnis kamen entgegen, dass bereits zeichengenau edierte Transkriptionen vorhanden waren und wiss. Hilfskräfte eingestellt werden konnten, die sowohl paläographische Kenntnisse als auch Interesse an digitalem Arbeiten mitbrachten.

Das Projekt wird ggf. demnächst fortgeführt und erweitert.