

Digitalisierung ausgewählter Libretti der Hamburger Gänsemarktoper von 1678-1730

PD Dr. Katrin Dennerlein

Dr. Christian Reul

Lücke in der deutschsprachigen Dramengeschichte von 1680-1730

Zwischen Christian Weise und Johann Christian Gottscheds Dramen gibt es fast keine Dramen, die den ästhetischen Ansprüchen und der aufklärerischen Fokussierung der Literaturhistoriker entsprechen. Das heißt, sie sind kein

- aufklärerisches
- original deutschsprachiges
- gedrucktes
- Sprechtheater

Verdacht : Erfundener, extemporiert Schwulst ohne Bedeutung.

Tatsächlich:

- gedruckte Libretti, davon auch ein beträchtlicher Anteil deutschsprachig (bzw. parallele ÜS)
- (häufig) vollausformulierte Sprechtheatertexte der deutschsprachigen Wanderbühne
- Schultheater

Lücke in den digitalisierten Dramenkorpora zwischen von 1680-1730

- Textgrid
- Gerdracor
- Volltextdigitalisierung durch die ÖNB und die StaBi München – theoretisch vorhanden, aber sehr fehlerhaftes OCR

Libretti der Hamburger Gänsemarktoper von 1678-1730

- Das wichtigste Korpus von deutschsprachigen Libretti aus diesem Zeitraum
- 270 Werke in ca. 2000 Textbüchern
- **95% der Libretti sind deutschsprachig, 55% sind Eigenproduktionen.**
- Biblische, historische und mythologisches Stoffe
- Jedes erhaltene Werk ist inzwischen in mind. einem Exemplar bilddigitalisiert.



Hamburger Gänsemarktoper

Das umfangreichste und vollständigste Corpus dramatischer Texte, das im 17. und 18. Jahrhundert für ein Theater geschrieben und gedruckt wurde.

Die Funktionen der Opern der Hamburger Gänsemarkt-Oper

- Repräsentation
- Vermittlung biblischer Inhalte
Erste Oper z.B. „(Adam und Eva) Der erschaffene/gefallene und
auffgerichtete Mensch.“ 1678 (Theile/Richter)
- Erbauung und Geschichtlichkeit
- Unterhaltung

Autor/Komponist	Titel	UA/ED
Bostel, Lucas von/ Franck, Johann Wolfgang	Der Glückliche Grosz-Vezier Cara Mustapha, Erster Theil	1686
Postel, Christian Heinrich/Keiser, Reinhard	Der bey dem allgemeinen Welt-Friede von dem Grossen Augustus geschlossene Tempel des JANUS	1698
Hotter/Keiser, Reinhard	Störtebecker und Jödge Michaels, 1. Teil	1701
Postel, Christian Heinrich/Keiser, Reinhard	Störtebecker und Jödge Michaels, 2. Teil	1701
Bressand, Friedrich Christian/Mattheson, Johann	Der Edelmüthige Porsenna	1702
Bressand, Friedrich Christian/ Mattheson, Johann	Der Gestürzte und wieder Erhöhte Nebucadnezar, König zu Babylon/Unter dem Großen Propheten Daniel	1704
Bressand, Christian/Keiser, Reinhard	Die Kleinmühtige Selbst-Mörderin Lucretia. Oder: Die Staats-Thorheit des Brutus	1705
Feind, Barthold/Keiser, Reinhard	Masagniello Furioso Die Neapolitanische Fischer-Empörung	1706
Feind, Barthold/Graupner, Christoph	Der Fall Des grossen Richters in Israel/Simson, Oder: Die abgekühlte Liebes-Rache der Debora	1709
Weidemann, Christian Heinrich/Keiser, Reinhard	Le Bon vivant, Oder die Leipziger Messe	1710
König, Johann Ulrich/Keiser, Reinhard	Die Oesterreichische Großmuth/Oder: Carolus V.	1712
König, Johann Ulrich/Keiser, Reinhard	Die wiederhergestelte Ruh/Oder Die gecrönte Tapferkeit des Heraclius	1712
König, Johann Ulrich/Keiser, Reinhard	Fredegunda	1715
Feind, Barthold/Händel, Georg Friedrich	Rinaldo	1715
Hoë, Johann Joachim/Keiser, Reinhard	Der die Vestung Siebenbürgisch-Weissenburg erobernde und über die Däcier triumphirende Kayser Trajanus	1717
Praetorius, Johann Philipp/Keiser, Reinhard	Der Hamburger Jahr-Marckt Oder der Glückliche Betrug	1725
Praetorius, Johann Philipp/Keiser, Reinhard	Die Hamburger Schlacht-Zeit/Oder Der Mißgelungene Betrug	1725
Praetorius, Johann Philipp/ Telemann, Georg Philipp	Calypso Oder Sieg der Weißheit Ueber Die Liebe	1727
Telemann, Georg Philipp / Telemann, Georg Philipp	Ernelinda 1730	

Ambrosiens höchstwunderbaren Safft,
Der die Unsterblichkeit verschafft,
Verschwendrisch einzufuchen.

Cal. Oeh' eilia wieder hin! warum verließ't du sie?

Proth. Dir Göttin Nachricht zu ertheilen,
Daß meine Mühe
Den Zweck noch nicht erreicht. Jezt ist er ganz allein,
Und Eucharis wird schon dich zu bedienen, eilen,

Cal. Um ihre Pein
Für die noch weiter zu verstecken.
Sie soll die Deines seyn,
Du kanst Calypsoen trösten.

(Allein vorherö will ich das Vergnügen schmecken;
Ihm noch eunuchen zu seuchen.)

Aria.
Eilt, Huld-Göttinnen mich zu schmücken,
Bewaffne mich, Venus, mit reizenden Blicken,
Die Augen welche mich entzünden,
Zu gleicher Flamme zu verbinden!
Doch sollte nicht der Falsche teuschen,
So laßt, ihr Furien, mich einen Beistand finden,
Ihn zu ermorden, zu zerfleischen.

Zweiter Auftrit.

(Ein Wald.)

Telemachus in einem Zelle schlaffend.

Mentor.

Ment. O Klüßelt - - und im Traum; - - ein schmeicheln des Gesicht,
So ihm vergnügte Stunden,
Und ein Elysich Feld verspricht,
Hält seinen Geist gebunden.
Verhöret Telemach! dein unvollkommene Sachen
Verblendet machen,

Ambrosiens höchstwunderbaren Safft.

Der die Unsterblichkeit verschafft,"

Verschwendrischeinzufuchen.-

Cal.. Geheilig wiederhin, warum verließ't du sie?-

Proth, Dir Göttin Nachricht zu ertheilen.

" Daß meine Mühe.....

Den Zweck noch nicht erreicht. Jezt ist er ganz allein,

Und Eucharis wird schöndlichzubedienen eilen,-

Um ihre Pein -----

- ürdirnoch weiterzuverstecken.....]

Cal. Ä In-

Dukanst Calypsoentrauen.----

(Alleinvorheröwillich das Vergnügen schmecken
Ihnenochmahlaufschauen)."

: 3 3. i: so Geyes ta)-

Aria. "I

„... TT:--

Eilt, Huld-Göttinnen mich zu schmücken.

- Bewaffne mich, Venus, mitreizenden Blick

Die Augen welchemichentzündend.. "

Ä Flamme zu verbinden!?

Doch sollte nicht der Falsche teuschen Ä

Solaßt, ihr Furienmicheinen Beistand finden

Ihnzuermordenjersehen j

- (Set ab.)

Zweyter Auftrit.- "

Ambrosiens höchst wunderbaren Safft/
 Der die Unsterblichkeit verschafft,
 Verschwendrich einzufuchencken.
 Cal. Geh' eilia wieder hin! warum verließ't du sie?
 Proth. Dir Göttin Nachricht zu ertheilen,
 Daß meine Müß
 Dem Zweck noch nicht erreicht. Jezt ist er ganz allein;
 Und Eucharis wird schon dich zu bedienen, eilen,
 Um ihre Pein
 Für dir noch weiter zu verstecken.
 Cal. Sie soll die deine seyn,
 Du kanst Calypsen trauen.
 (Allein vorhero will ich das Vergnügen schmecken;
 Jhn noch einmahl zu schauen.)

Aria.

Eilt, Huld-Göttinnen mich zu schmücken,
 Bewaffne mich, Venus, mit reizenden Blicken,
 Die Augen welche mich entzünden,
 Zu gleicher Flamme zu verbinden!
 Doch sollte mich der Falsche teuschen,
 So laßt, ihr Furien, mich einen Beistand finden,
 Jhn zu ermorden, zu zerfleischen.

(Geh ab.)

Zweyter Auftritt.

(Ein Wald.)

Telemachus in einem Zelle schlaffend.

Mentor.

Ment. R lächelt - - und im Traum; - - ein schmeichelndes Gesicht,
 So ihm vergnügte Stunden,
 Und ein Elyfisch Feld verspricht,
 Hält seinen Geist gebunden.
 Bethörter Telemach! dem unvollkommne Sachen
 Verblindet machen,

#809 Ambrosiens höchst wunderbaren Safft/
 #810 Der die Unsterblichkeit verschafft,
 #811 Verschwendrich einzufuchencken.
 #812 Geh' eilia wieder hin! warum verließ't du sie?

#813 Proth. Dir Göttin Nachricht zu ertheilen,
 #814 Daß meine Müß
 #815 Dem Zweck noch nicht erreicht. Jezt ist er gantz allein,
 #816 Und Eucharis wird schon dich zu bedienen, eilen,
 #817 Um ihre Pein
 #818 Für dir noch weiter zu verstecken.

#819 Cal. Sie soll die deine seyn,
 #820 Du kanst Calypsen trauen.
 #821 (Allein vorhero will ich das Vergnügen schmecken;
 #822 Jhn noch einmahl zu schauen.)

#823 (Beyde gehet ab.)

#824 Aria.

#825 Eilt, Huld-Göttinnen mich zu schmücken/
 #826 Bewaffne mich, Venus, mit reizenden Blicken,
 #827 Die Augen welche mich entzünden/
 #828 Zu gleicher Flamme zu verbinden!
 #829 Doch sollte mich der Falsche teuschen,
 #830 So laßt, ihr Furien/ mich einen Beistand finden/
 #831 Jhn zu ermorden/ zu zerfleischen.

#832 (Geh ab.)

#833 Zweyter Auftritt.

#834 (Ein Wald.)

#835 Telemachus #836 in einem Zelle schlaffend. #837 Mentor.

#838 Ment. R lächelt - - und im Traum; - - ein schmeichelndes Gesicht/
 #839 So ihm vergnügte Stunden/
 #840 Und ein Elyfisch Feld verspricht,
 #841 Hält seinen Geist gebunden.
 #842 Bethörter Telemach! dem unvollkommne Sachen
 #843 Verblindet machen,

OCR – Grundlagen: Workflow und Modelle



- Hauptkomponenten: Vorverarbeitung, Segmentierung, OCR, Nachkorrektur
- Moderne OCR-Ansätze arbeiten auf Zeilen- und nicht mehr auf Zeichenbasis
- Sog. *Modelle* extrahieren Text aus Textzeilenbildern
 - Modelle müssen trainiert werden (Neuronale Netze)
 - Trainingsdaten bestehen aus Zeilenbild-Text-Paaren

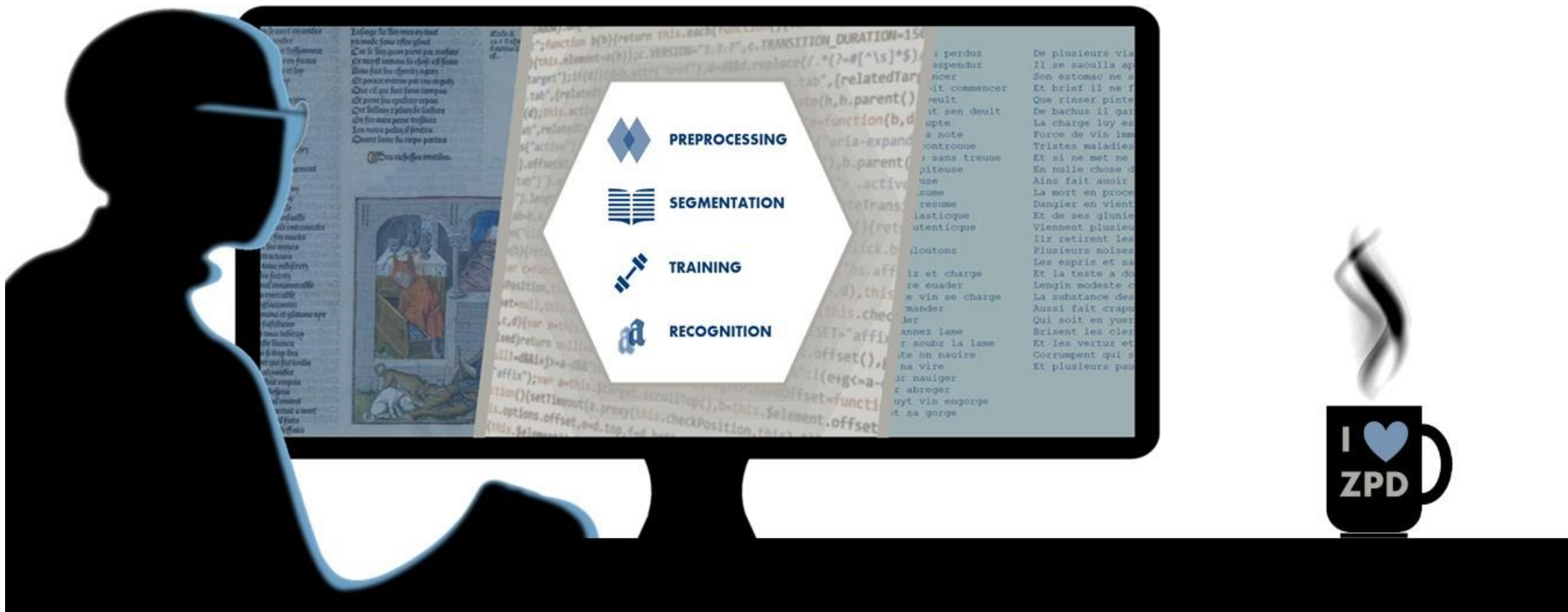
Er wird eifrig gesammelt.
Er wird eifrig gesammelt.

OCR4all – Motivation

- Bestehende Open Source OCR Software ist mächtig, kann nicht-technische Nutzer jedoch schnell überfordern:
 - Installation nicht unbedingt trivial
 - Keine grafische Benutzeroberfläche, sondern ungewohnte Kommandozeile
 - ...
- Die Idee hinter OCR4all:
 - Verständlich und anwendbar auch für nicht-technische Nutzer
 - Nutzung von Docker/VirtualBox → Einfache Installation, Unabhängigkeit vom Betriebssystem
 - Basiert auf unterschiedlichen Open Source OCR Tools
- Ursprünglich im Rahmen des Editionsprojekts [Narragonien digital](#) entwickelt, für die interaktive und hochqualitative Erfassung sehr alter Drucke

OCR4all – Komponenten

- OCRopus
 - <https://github.com/ocropus/ocropy>
 - Vorverarbeitung (Binarisierung, Deskewing)
 - Automatische (Zeilen)Segmentierung
- LAREX
 - <https://github.com/OCR4all/LAREX>
 - Ursprünglich zur interaktiven Regionensegmentierung inklusive semantischer Auszeichnung entwickelt
 - Mittlerweile umfangreiches Korrekturtool, u. a. für
 - Regionen- und Zeilenkoordinaten
 - Semantische Typisierung und Lesereihenfolge
 - Text (Ground Truth Erstellung für Training!)
- Calamari
 - <https://github.com/Calamari-OCR>
 - Texterkennung
 - Modelltraining
 - Evaluation



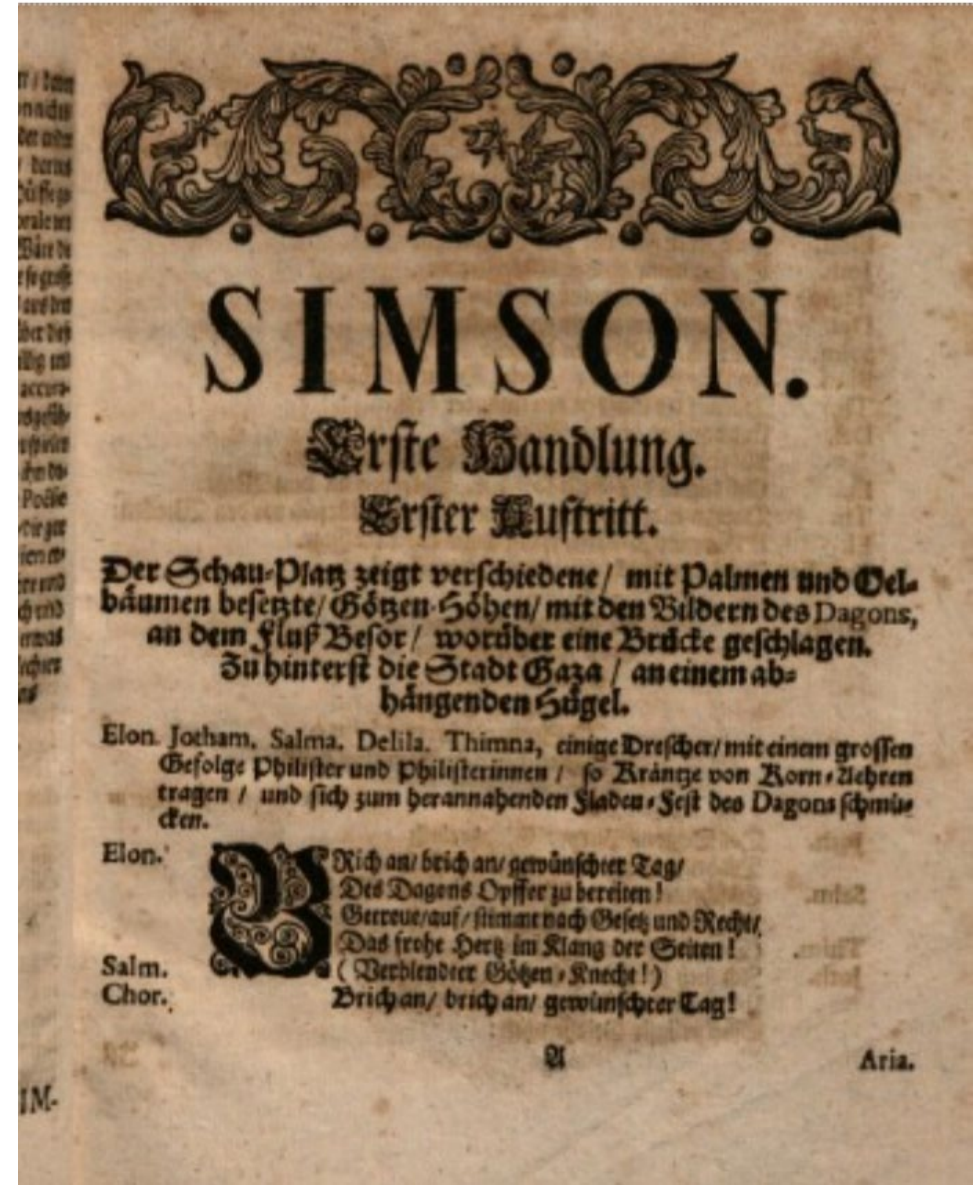
www.ocr4all.de

Live Demo!



Libretti-Workflow

- Festlegung dessen, was erfasst werden soll (Unterschiede zu einer Edition)
- Preprocessing (automatisch): Eingabeformat (Bild oder pdf), Umwandlung in png
- Segmentierung und Zuweisung einer Lesereihenfolge zu den Segmenten (händisch)
- Zeilensegmentierung (automatisch, händische Korrektur)
- OCR-Korrektur Zeile für Zeile (iteratives, automatisches Training, manuelle Korrektur)
- Zuweisung von TEI-tags zu Segmenten und Verschachtelung von TEI-tags (automatisiert, händische Nachbearbeitung)



Segmentierung (semantische Auszeichnung)

- Text (Haupttext)
- Text (Vorrede)
- Sprechername
- Sprechername (>1 Sprecher) z.B. „Chor“, „Alle“
- Nebentext
- Fußnote
- Überschrift (Akt)
- Überschrift (Szene)
- Überschrift (Gesang)
- Strophennummerierung
- Überschriften (Sonstige)
- Personenverzeichnis (Figurenname)
- Personenverzeichnis (Figurenbeschreibung)
- Bild, Holzschnitt, Diagramm, Tabelle, Initiale, Zierinitiale, Formel...

Beispiel für Segmentierung mit Zuweisung
der Reihenfolge

ert / Daten
ion nicht
ider ande
/ Darau
Hüfte ge
iorale ben
Wäre die
ne so gro
s aus dem
über die
villig und
s accura
tungsgefu
auspichen
d ihm dar
ie Poëtie
die gar
nien es
hre und
ich und
herwad
hlehren
ind



SIMSON.

Erste Handlung.

Erster Auftritt.

Der Schan-Platz zeigt verschiedene / mit Palmen und Oel-
bäumen besetzte / Götzen-Höhen / mit den Bildern des Dagon,
an dem Fluß Besor / wor³ er eine Brücke geschlagen.
Zu hinterst die Stadt Gaza / an einem ab-
hängenden Hügel.

Elon. Jotham. Sal⁴. Delila. Thimna, ei⁵ e Di⁶ ber / mit ein⁷ grossen
Gefolge Philister und Philisterinnen / so Kränze von Korn, Aehren
tragen / und sich zum herannahen⁸ in Glan, Seht des Dagon schmü-
cken.

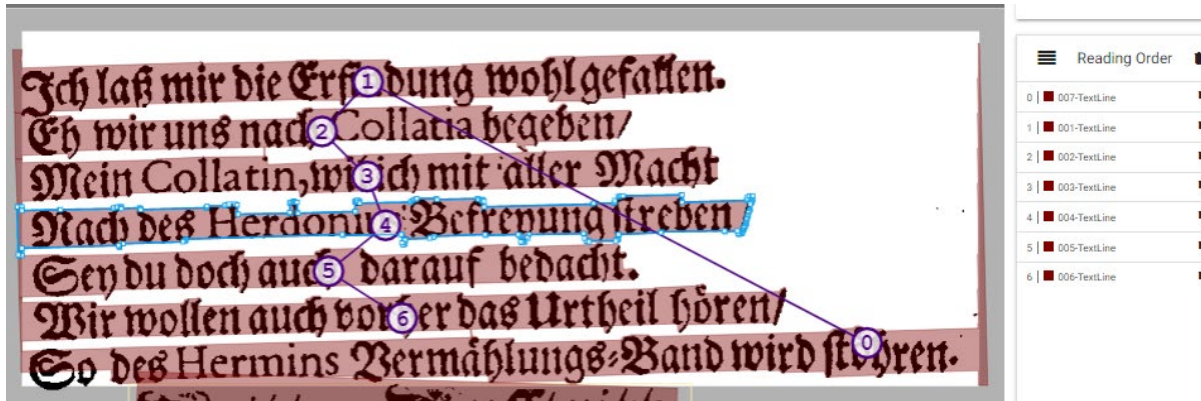
Elon.



Salm.
Chor.

Rich an / brich an / gewünschter Tag /
Des Dagon's Opffer zu bereiten !
Gerue / auf / stimme / ach Befehl und Recht !
Das frohe Herz im Klang der Seiten !
(Verblendter Götzen, Knecht !)
Brich an / brich an / gewünschter Tag !

Zeilensegmentierung



Jede Zeile wird als einzelne Zeile erkannt
ABER: Die letzte Zeile wurde fälschlicherweise als erste Zeile erkannt.



Zwei Zeilen werden fälschlicherweise als eine zusammengehörige erkannt



Hier wurde der Text nur unvollständig erkannt. Außerdem wurde eine Zeile als zwei Zeilen erkannt

Recognition, Korrektur, und Ground Truth Production

Händische Korrektur der automatischen
Erkennung

Training von vorhandenen Modellen zur
Erkennung von Frakturschrift. Ergebnis:
„c1_libretti_bin-nrm0–4“

Von 3,5% zu Beginn zu 1,5% Fehlerquote

Eine edle Römerin / vermählt an Collatinus.
Eine edle Römerin / vermählt an Collatinus.

Gemahlin des Turnus.
Gemahlin des Turnus.

Brutus junge Tochter / versprochen an Herminius und
Brutus junge Tochter / versprochen an Herminius und

nachmahls verliebt in Sextus.
nachmahls verliebt in Sextus.

Tarquinius Collatinus.
Larquinius Collatinus.

Obrister über die Cavallerie, und Ehgemahl
Obrifter über die Cavallerie, und Ehgemahl

Actus

Händische Korrektur der Recognition
Zeile für Zeile

Postcorrection

- An dieser Stelle können die verschiedenen Schritte Segmente, Linien und Text im LAREX nochmals manuell angepasst werden
- Zu beachten ist, dass eine Bearbeitung der höheren Ebene eine Nachbearbeitung auf niedrigerer Ebene nach sich zieht. Wird also zum Beispiel ein Segment gelöscht und neu gezogen, müssen auch die Linien und der Text neu gezogen bzw. erstellt werden.

Automatische Umwandlung in TEI

- TEI: semantische Auszeichnung von Textteilen (Akt, Szene, Sprecherbezeichnung)
- Regelbasiertes Mapping anhand semantischer Auszeichnung und Zeilenkoordinaten
- Manuelle Nachkorrektur, unterstützt durch Anzeige unsicherer Stellen:
 - Schmuckinitialien
 - Spezielle Sprecherzuordnungen (siehe rechts)
 - Auflösung bestimmter Sprecherbezeichnungen (à 2, Alle usw.)

The image shows a screenshot of a TEI XML editor interface. It displays a musical score with German lyrics. The text is annotated with TEI tags, which are highlighted with colored boxes. A red horizontal line is at the top. The annotations include:

- A green box containing "Aria à 2." above the first line of text.
- A blue box containing "Tel." above the first line of text.
- A blue box containing "Tel." above the second line of text.
- A blue box containing "à 2." above the third line of text.
- A blue box containing "Euch." above the fourth line of text.
- A red box containing "Euch." above the first line of text.
- A red box containing "Euch." above the second line of text.
- A red box containing "Wehrtes Leben!" above the first line of text.
- A red box containing "Bei den Blicken!" above the second line of text.

The lyrics are:

Schönster Engel!
Bei den Küssen,
Bricht, für freudigem Entzücken,
Fast das Herz in dieser Brust.
Kein Verändern soll uns trennen,
Weil mir solches unbekannt.

Beispiel für TEI-Auszeichnung

- Vgl. Dokument „Beispiele_Seiten.pdf“

Diskussion/Bewertung

- Typische Fehler/Probleme
- Zeitaufwand:
 - Preprocessing: Umwandlung in .png (halbe Stunde pro Libretto)
 - Segmentierung, Reihenfolge: für 70 Seiten bis zu fünf Stunden
 - Zeilensegmentierung (automatisch, manuelle Korrektur): vier Stunden automatische Umwandlung, 1-2 Stunden Fehlerkorrektur
 - OCR-Korrektur Zeile für Zeile (automatisch, manuelle Korrektur), 10 Minuten für die automatische Recognition, 2-4 Stunden für die Korrektur Zeile für Zeile (=ground truth production), 2 Stunden für händische Nachkorrektur
 - Zuweisung von TEI-tags zu Segmenten und Verschachtelung von TEI-tags (automatisiert, händische Nachbearbeitung) – Zeitaufwand momentan noch unklar
- Zeichenfehlerrate: mittlerweile bei 1,2%-1,5%
- Methodischer Mehrwert gegenüber Double-Keying
Für Double-Keying wäre das Festlegen und Vermitteln von Richtlinien sehr anspruchsvoll; keinerlei Weiterentwicklung von Techniken/Synergien

Vielen Dank für Ihre Aufmerksamkeit!