

TUSTEP-gestützte Digitalisierung deutschsprachiger Poetiken: Ein (retrospektiver) Blick auf das Projekt *ePoetics*

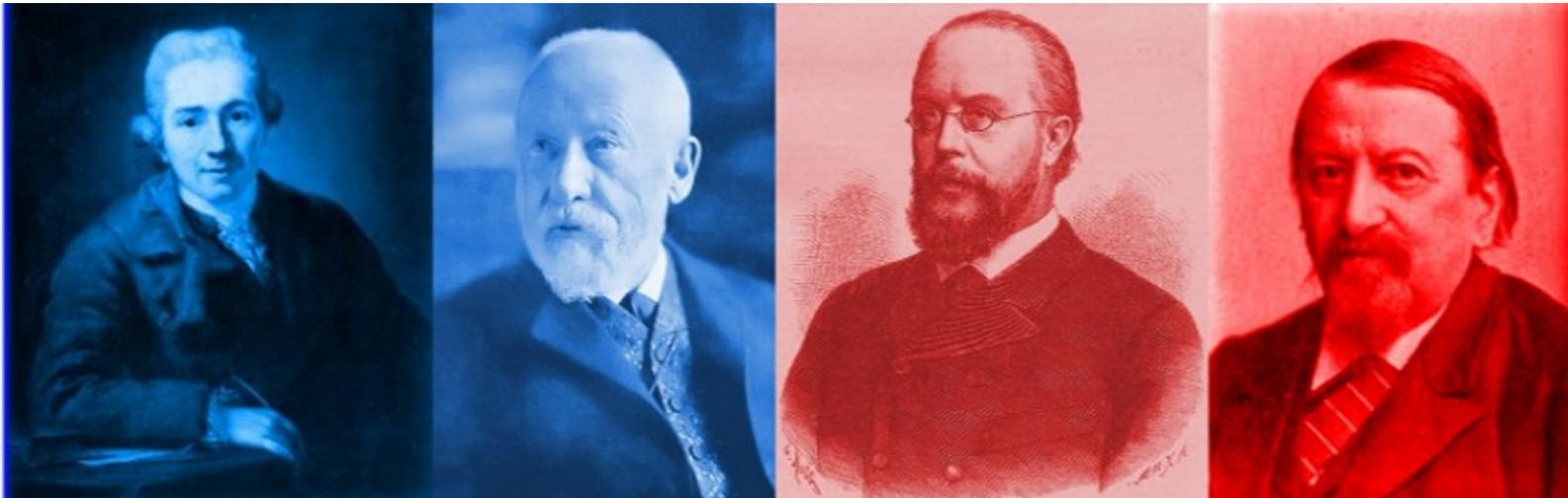


TECHNISCHE
UNIVERSITÄT
DARMSTADT

David Glück

(weitere Beteiligte: Dr. Michael Bender, Constanze Hahn, Hans-
Werner Bartz)

Universität Stuttgart



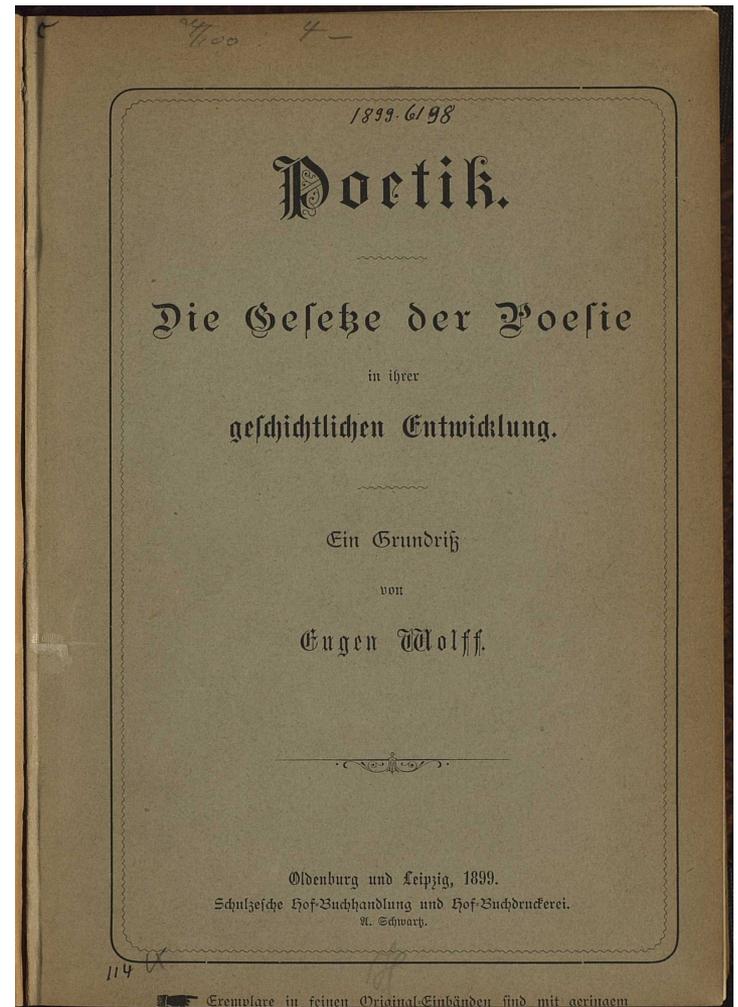
TECHNISCHE
UNIVERSITÄT
DARMSTADT



„ePoetics – Korpuserschließung und Visualisierung deutschsprachiger Poetiken (1770-1960)“

- Kooperationsprojekt der Universität Stuttgart (Germanistik, Informatik, Maschinelle Sprachverarbeitung) und der TU Darmstadt (Sprach- und Literaturwissenschaft)
- Laufzeit: 01.02.2013 - 31.01.2016, gefördert vom Bundesministerium für Bildung und Forschung
- Ziele: Digitalisierung ausgewählter deutschsprachiger Poetiken und Entwicklung computergestützter Methoden und Werkzeuge zur Analyse der Poetiken

- Poetiken: "Lehren von der Dichtkunst"; theoretisch-praktische Reflexionen über literarische Texte, Stilmittel, Prinzipien des dichterischen Schaffens etc.
- Korpus von 20 ausgewählten Poetiken von 1770 bis 1960 aus dem deutschen Sprachraum: philologisch einflussreiche, deskriptiv-wissenschaftliche Poetiken
→ Vorläufer moderner Literaturtheorie





- Verfügbarmachung der Texte in digitalen Textarchiven (*TextGrid Repository* und *Deutsches Textarchiv*) in annotierter Form:
 - strukturelle Annotation (Gliederungsebenen und -einheiten der Texte)
 - literaturwissenschaftlich-interpretative Annotation: literarische Konzepte und ihre Behandlung durch den jeweiligen Autor: Begriffsverständnis, zitierte Primär- und Sekundärliteratur, ... → Konzept der Metapher
- Entwicklung von computerlinguistischen Tools und Visualisierungstools zur Analyse
- Pilotstudie: Analyse der Poetiken durch Kombination aus literaturwissenschaftlich-hermeneutischen und informatischen Verfahren („Distant Reading“/„Algorithmic Criticism“)

ePoetics: Digitalisierung (I)



- Texte bisher kaum digitalisiert
 - kompletter Digitalisierungsprozess erforderlich
- Anfang: hochauflösende Scans der Erstausgaben
- Zielformat: XML, im Standard der *Text Encoding Initiative* (TEI)
 - gängiges Format, philologisch reflektiertes "Markup"
 - einfache Integration in bestehende Infrastrukturen wie *TextGrid* und *DTA* (Weiterverarbeitung in HTML)

```
<pb facs="#f0005" n="EAL:b"/><div n="1">
<lb n="pwo_000.001"/><head><hi rendition="#c"><hi
rendition="#g">Poetik</hi>. </hi>
<lb n="pwo_000.002"/><hi
rendition="#c"><figure/></hi></head>
<lb n="pwo_000.003"/><head><hi rendition="#c"><hi
rendition="#g">Die Gesetze der Poesie</hi>
<lb n="pwo_000.004"/>in ihrer
<lb n="pwo_000.005"/>geschichtlichen
Entwicklung.</hi></head>
<lb n="pwo_000.006"/><p><hi
rendition="#c"><figure/></hi></p>
<lb n="pwo_000.007"/><p><hi rendition="#c">Ein
Grundriß
<lb n="pwo_000.008"/>von
<lb n="pwo_000.009"/><hi rendition="#g">Eugen
Wolff</hi>. </hi></p>
<lb n="pwo_000.010"/><p><hi
rendition="#c"><figure/></hi></p>
<lb n="pwo_000.011"/><p><hi rendition="#c">Oldenburg
und Leipzig, 1899.
<lb n="pwo_000.012"/>Schulzesche Hof-Buchhandlung und
Hof-Buchdruckerei.
<lb n="pwo_000.013"/>A. Schwartz.
<lb n="pwo_000.014"/><figure/> Exemplare in feinen
Original-Einbänden sind mit geringem</hi></p>
<pb facs="#f0006" n="EAL:a"/>
<pb facs="#f0007" n="RI"/></div><div n="1">
```



Digitalisierungsablauf:

- 1.) Transkription (+ Vorannotation)
- 2.) literaturwiss. Annotation (Beispiel-Konzept: Metapher)
- 3.) XML-Auszeichnung – 2 Phasen:
 - Umwandlung der Vorannotationen in den Transkriptionen in XML/TEI-Tags mittels TUSTEP
 - Nachbearbeitung und Validierung der mit TUSTEP bearbeiteten Dateien (v.a. manuell in *oXygen*)
- 4.) Entwicklung eines geeigneten XSLT zur Umwandlung des XML in HTML

Transkription (I)



- Von externem Dienstleister übernommen: Double Keying-Verfahren
- Anreicherung der transkribierten Texte mit grundlegendem Markup:
 - typografisch: Schriftart/-größe, Textausrichtung, Layout
 - strukturell: Texteinheiten (z.B. Paragraphen, Seitenumbrüche, Fußnoten, ...)

```
<pb n="E1"/>
<L PWO 001.001/><Z><W+3>^#S+Einleitung^#S-./W+3)../Z)
<L PWO 001.002/><Z>^#(IMAGE)</Z)
<L PWO 001.003/><Z><W+2>Begriff und Methoden der Poetik
./W+2)../Z)
<L PWO 001.004/><Z>§ 1.
<L PWO 001.005/><W+1>Begriff der Poetik./W+1)../Z)
<L PWO 001.006/><P> Poetik ist die Wissenschaft von
den Gesetzen der Poesie.</P)
<L PWO 001.007/><P> Eine Wissenschaft ist die
systematisch geordnete Summe dessen,
<L PWO 001.008/>was wir über ein Gebiet wissen.
Folglich ist Material der Poetik,
<L PWO 001.009/>aus dem sie ihre Gesetze ableitet,
die systematisch geordnete Summe
<L PWO 001.010/>dessen, was wir über die Poesie
wissen; d. h. Gegenstand der Poetik
<L PWO 001.011/>ist die systematisch geordnete
Geschichte der Weltpoesie.</P)
<L PWO 001.012/><P> Als Voraussetzungen für eine
wissenschaftliche Poetik ergeben sich:</P)
<L PWO 001.013/><E><F><A>a./A)) Ausgehen von der
Weltpoesiegeschichte,</P></E)
<L PWO 001.014/><E><negEZ><A>b./A)) Ordnung der
Weltpoesiegeschichte nach einheitlichen
Gesichtspunkten.</negEZ></E)
```

Transkription (II)



- Kodierung in Unicode (UTF8)
- Benutzung von TUSTEP-Steuerzeichen/-Kodierungen, v.a. bei
 - typografischen Elementen
 - nicht-lateinischen Zeichensystemen (v.a. Altgriechisch, Hebräisch)
 - Sonderzeichen, z.B. metrische Betonungszeichen:
#(MKS) → ∪ ; #(MLS) → —
 - Diakritika, v.a. für komplexere Überstellungen: rēh%)=sē → rēh≐sē

⟨P⟩ Frühere Erklärungen des Begriffs Rhythmus stehen mit unserer Definition nur teilweise in Widerspruch. ^#S+Platon^#S- definiert Rhythmus als Ordnung der Bewegung. ^#G+τῆ τῆς κινήσεως τάξει ρυθμός ὄνομα εἶη^#G- ⟨A⟩Platon de legg. II⟨/A⟩.) ^#S+Aristides Quintil.^#S- bezeichnet ihn als ^#S+Zusammensetzung geordneter Zeitlängen^#S- und meint, der Rhythmus lasse sich erkennen

Textauszeichnung: TUSTEP (I)



TUSTEP als Texteditor:

- Zeigeanweisungen (zn, zu):
Identifikation der
Vorannotationen (z.B. zn,,,|#?
+|#?-|^<*>|^</*>|)
- Austauschweisungen (a):
Umwandlung der
Vorannotationen in valide
XML/TEI-Tags (z.B. a,,,|
#S+*#S-|<hi
rendition="#g">{+4=}</hi>|)

```
1.530 <L_P1C_XXV.016/>der #S+rationalen Psycholo
1.531 <L_P1C_XXV.017/>zur Deduction des #S+Schön
1.532 <L_P1C_XXV.018/>^<negEZ^>§. 4. 5. Vorläuf.
1.534 <L_P1C_XXV.020/>^<negEZ^>§. 6. #S+Empirisi
1.535 <L_P1C_XXV.021/>und #S+niederer#S- Schöne
1.536 <L_P1C_XXV.022/>^<negEZ^>§. 7. #S+Unterga
1.537 <L_P1C_XXV.023/>aufgestellt. Das #S+Heftig
1.538 <L_P1C_XXV.024/>das #S+Große,#S- das #S+E
1.539 <L_P1C_XXV.025/>#S+Hauptelemente#S- des #
1.540 <L_P1C_XXV.026/>^<E^>^<negEZ^>#S+Anmerk.#
1.541 <L_P1C_XXV.027/>Nationen in Ansehung dies
1.543 <L_P1C_XXVI.001/>#S+mente. Scheidung#S- u
1.545 <L_P1C_XXVI.003/>Beispiele des #S+Starken
1.548 <L_P1C_XXVI.006/>(das #S+tragische#S- patl
1.549 <L_P1C_XXVI.007/>#S+Großen#S- S. 105. ers
1.551 <L_P1C_XXVI.009/>S. 120. #S+Beispiele#S-
1.555 <L_P1C_XXVI.013/>himmlisch erhabene S. 13
1.556 <L_P1C_XXVI.014/>^<E^>^<negEZ^>#S+Anmerk.
1.557 <L_P1C_XXVI.015/>des #S+höheren Schönen#S-
1.559 <L_P1C_XXVI.017/>und mit der Gemüthsstimme
1.560 <L_P1C_XXVI.018/>wunderbare, romantische
1.561 <L_P1C_XXVI.019/>^<E^>^<negEZ^>#S+Anmerk.
```

699 Zchflg. gefunden

258 Sätze gefunden

Textauszeichnung: TUSTEP (II)

TUSTEP als Skript-Werkzeug

- Zusammenführung von Fußnoten per Skript: Fußnotenreferenz im Haupttext, separate Fußnotendatei, Zusammenf. per Einfüge-Programm

nicht mehrere, die wir schwerlich alle
mögen aufzählen können. In einer von
Geßner übersetzten Erzählung, oder vielmehr
Unterredung, <i>Diderots</i> <fn id="PEN 440 *"/>) werden
mehrere kleine Facta, zum Theil nur unausgeführte
Situationen nacheinander hingeworfen (+{FN}+)

<FN id="PEN 440 *">

<L PEN 440.020/>) Man s. unsers Verfassers Urtheil darüber: Bd 4,
<L PEN 440.021/>S. 195 folg.</FN>

<lb n="pen_440.017"/>Unterredung, <hi
rendition="#i">Diderots</hi>

<note place="foot" xml:id="pen_440_1" n="*">
<lb n="pen_440.020"/> Man s. unsers Verfassers Urtheil
darüber: Bd 4, <lb n="pen_440.021"/>S. 195 folg.</note>
werden

<lb n="pen_440.018"/> mehrere kleine Facta, zum Theil nur

```
#ei,tx0,tx1,kurz,+,*,kuer=fn
ber      731.52-731.52
* Quelle
akd      |<fn id="|<note place="foot" n="|<note type="footnote" n="|
ekd      |"/>|"/>|
* Kuerzeldatei
akk      |<fn id="|
ekk      |">|

kkz      1 1 1
mkl      90
nae      |yyy|
```

Textauszeichnung: TUSTEP (III)



TUSTEP als Skript-Werkzeug

- Integration der häufigsten Austauschweisungen in das Skript
→ weitgehende Automatisierung der XML-Auszeichnung

```
#ko,tx2,tx3,+,+,*  
xx      | ^<{0-1}/p^>|<{+2=}p>|  
xx      | ^<{0-1}/shi^>|<{+2=}lg>|  
xx      | ^<z^>|<hi rendition="#c">|  
xx      | ^</z^>|</hi>|  
xx      | ^<a{0-2}?^>|<hi rendition="#aq">|  
xx      | ^</a{0-2}?^>|</hi>|  
xx      | ^<{0-1}/w??^>||  
xx      | ^<e^>|<hi rendition="#et">|  
xx      | ^</e^>|</hi>|  
xx      | ^<duolan 2^>|<cb type="start"/>|  
xx      | ^</duolan 2^>|<cb type="end"/>|  
xx      | @@|<cb/>|  
xx      | #g+|<foreign xml:lang="grc">#g+|  
xx      | #g-|#g-</foreign>|  
xx      | #S+|<hi rendition="#g">|  
xx      | #S-|</hi>|
```

Textauszeichnung: Nachbearbeitung

Hauptsächlich mit *oXygen* (XML-Editor):

- Erweiterung der Annotation, v.a. strukturelle Annotation über größere Textteile, z.B. Kapitel, Tabellen, Listen, Inhaltsverzeichnisse etc.
- XML-Validierung (nach TEI-Schema des *Deutschen Textarchivs*) → korrekte Elemente, Attribute, XML-Struktur etc.
 - "Feinheiten" der Textauszeichnung
- Verknüpfung mit (selbstentwickeltem) XSLT zur Transformation des XML in finale HTML-Versionen

Besondere Auffälligkeiten bei der Benutzung von TUSTEP

- enorme Performanz: Texteditor-Funktionen und Skripte
- Vielfalt der Funktionen: (mächtige) Texteditorfunktionen und Skripte in *einer* Programmumgebung
- Flexibles Zeichenkodierungssystem: Kombination von „normalem“ Zeichensatz (z.B. UTF-8) und TUSTEP-Steuerzeichen zur Kodierung von Zeichen (wie Diakritika)
- Schwierigkeiten:
 - Kodierung: teilweise „kryptisch“ – erfordert Einarbeitung; Repräsentation nicht-lateinischer Schriftsysteme mittels lateinischer Schrift → umständliche Korrektur
 - Zeige-/Austauschanweisungen nicht über Datensätze hinweg möglich
 - keine XML-Validierung in TUSTEP (?)

Fazit: TUSTEP in ePoetics (II)

Nutzen von TUSTEP in ePoetics:

- passendes Tool für die Textauszeichnung eines größeren Korpus für literaturwissenschaftliche Zwecke
- bei der Arbeit in ePoetics reichten wenige Funktionalitäten aus (einfache Editoroperationen + Skripte)
- v.a. geeignet als *ein* (zentraler) Baustein des Digitalisierungsworkflows (Posttranskription) auf dem Weg zur digitalen Publikation – in Kombination mit anderen Tools (wie oXygen)

Deutsches Textarchiv. Berlin-Brandenburgische Akademie der Wissenschaften 2007-2016.

<<http://deustchestextarchiv.de/>>

ePoetics: Korpuserschließung und Visualisierung deutschsprachiger Poetiken (1770-1960) für den „Algorithmic criticism“. <<http://www.epoetics.de/>>

Moretti, Franco: Distant Reading. London, New York 2013.

Ramsay, Stephen: Algorithmic Criticism. In: Siemens, Ray; Schreibman, Susan (Hgg.): A Companion to Digital Literary Studies. Chichester, Malden, Oxford 2013. S. 477-491.

Richter, Sandra: A History of Poetics. German Scholarly Aesthetics and Poetics in International Context, 1770-1960. Berlin, New York 2010.

TEI: Text Encoding Initiative. <<http://www.tei-c.org/>>

TextGrid Repository. TextGrid 2016. <<https://textgridrep.org/>>