



# Das Deutsche Textarchiv und die korpusbasierte Analyse komplexer Textmuster

Matthias Boenig (OCR-D) und Linda Kirsten (t.evo)

boenig@bbaw.de

linda.kirsten@bbaw.de

<http://www.upb.de/tevo>

<http://www.deutschestextarchiv.de>

# Gliederung

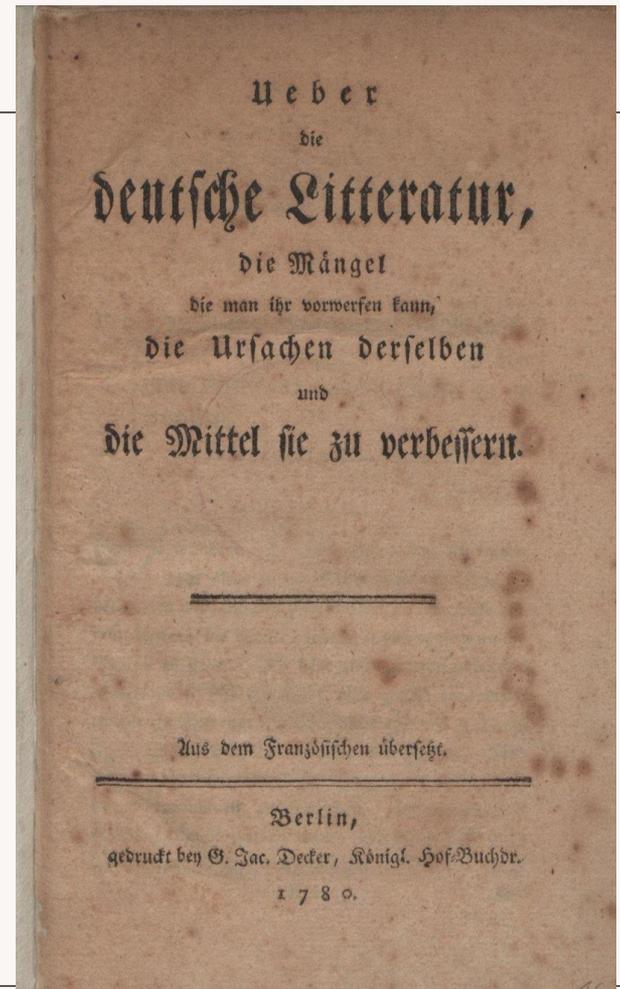
---

- Motivation
- Zahlen des Textarchivs
- Korpora des Textarchivs
- Zugänge zum Textarchiv

# Motivation

“Wir besitzen noch keine von der ganzen Nation gebilligte Sammlung, in der man alle Worte und Redensarten fände, nach denen man die Reinigkeit der Sprache sicher beurtheilen könnte.”

Quelle: Friedrich II., König von Preußen: Über die deutsche Literatur. Übers. v. Christian Konrad Wilhelm Dohm. Berlin, 1780, S. [1]. In: Deutsches Textarchiv <[https://www.deutschestextarchiv.de/friedrich\\_literatur\\_1780/7](https://www.deutschestextarchiv.de/friedrich_literatur_1780/7)>, abgerufen am 09.11.2021



Nachweis: Berlin SBB-PK, Bibl. Diez oct. 8354

# Zahlen des Textarchivs

	Werke	Wortformen	Zeichen
DTA-Kernkorpus	1 468	129 Mio.	904 Mio.
DTA-Erweiterungen	5 014	189 Mio.	1,3 Mrd.

Quelle: Deutsches Textarchiv – Projektüberblick, Online:

<https://www.deutschestextarchiv.de/doku/ueberblick>, abgerufen am 09.11.2021

was dem einen recht ist, ist dem andern billig

“In Deutschland, wie billig für deutsche Verhältnisse, hatten wir keine große Nationalitäts-Poesie; desto ungestümer brach die Individualitäts-Periode an: — das war die Sturm- und Drang-Periode.”

Quelle: Kürnberger, Ferdinand: Der Amerika-Müde. Frankfurt (Main), 1855, S. 147. In: Deutsches Textarchiv  
<[https://www.deutschestextarchiv.de/kuernberger\\_amerikamuede\\_1855](https://www.deutschestextarchiv.de/kuernberger_amerikamuede_1855)  
am 09.11.2021.

<https://www.deutschestextarchiv.de>

Angst, er springt auf, rennt was er kann aus dem Bereich des Dichters — und zu glücklich, wenn nicht der Nachklang noch fortführt, ihm seine innere Stimme zu verwirren! Das ungefähr war mein poetischer Erfindungsstand. Ich machte eiligst eine Skizze von meiner Melodie, warf sie in Gotta's Briefschalter und rannte auf und davon nach Amerika. In der Stille des Hinterwalds will ich sehen, ob ich die Skizze ausführe. — Moorfeld fuhr fort: Ich sagte zuvor: unsere ganze Poesie drückte nicht aus, was sie sollte und wollte: das befreimdete Sie. Ich bin Ihnen, wie es scheint, eine Erklärung darüber schuldig?

Es interessiert mich, sie zu hören, antwortete Benthal.

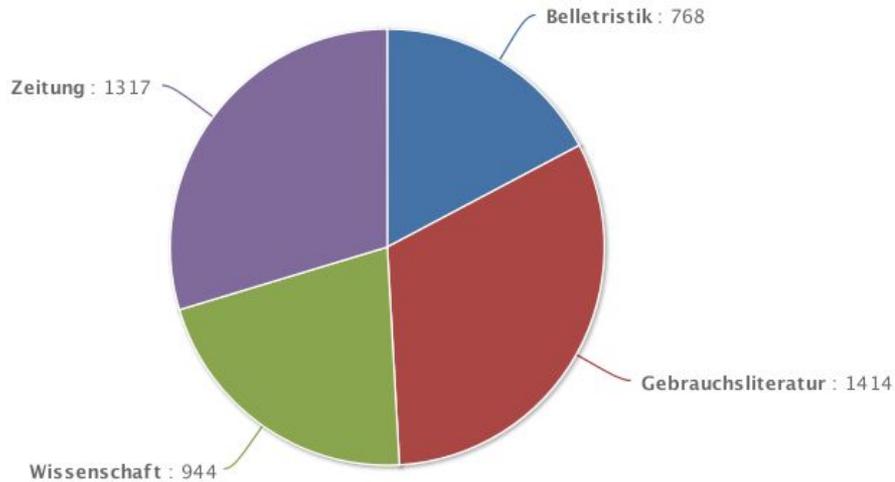
Ich meine es so, sagte Moorfeld: die ganze Literaturgeschichte zerfällt mir in zwei Perioden; die eine zähle ich von Homer bis Racine, die zweite von Racine bis in unbekannte Zeiten. Diese Perioden mögen Ihnen wunderbar dünken; in der ersten stehen z. B. die großen Gegensätze von antik und romantisch, christlich und heidnisch unberücksichtigt neben einander, — aber ich finde ein Merkmal der Gleichartigkeit für sie: den Ausdruck des nationalen Inhalts. Homer singt seine Griechen, Cervantes seine Spanier, Camoens seine Portugiesen, Shakespear seine Engländer, bis herauf zu Racine, welcher seine Franzosen singt. Das ist das einheitliche Moment dieser Periode — die Poesie der Nationalität. Nach Racine folgt eine andere Periode — die Poesie der Individualität. Recht schlagend für diese Eintheilung mag ich zwei Engländer nennen — Shakes-

und neue Zeit meines Begriffes. — In Deutschland, wie billig für deutsche Verhältnisse, hatten wir keine große Nationalitäts-Poesie; desto ungestümer brach die Individualitäts-Periode an: — das war die Sturm- und Drang-Periode. Man hat von einem Abschluß dieser

Periode durch Schiller und Goethe gesprochen. Aber Sie sehen wohl, wie lächerlich das ist. Haben wir denn bis auf diese heutige Stunde schon einen andern Inhalt gewonnen, als den der Sturm- und Drang-Periode — unser armes drangvolles Ich? Oder ist dieses Ich so

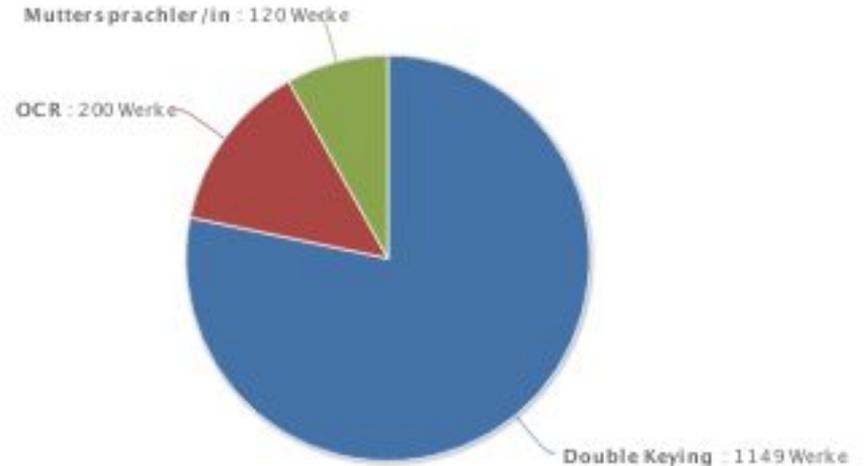
# Textgrundlagen

Im DTA verfügbare Werke (Kernkorpus + DTAE)  
nach Genre



Werke nach Digitalisierungsmethode

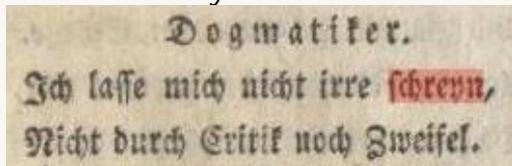
Hinweis: In dieser Statistik sind nur Werke aus dem DTA-Kernkorpus erfasst.



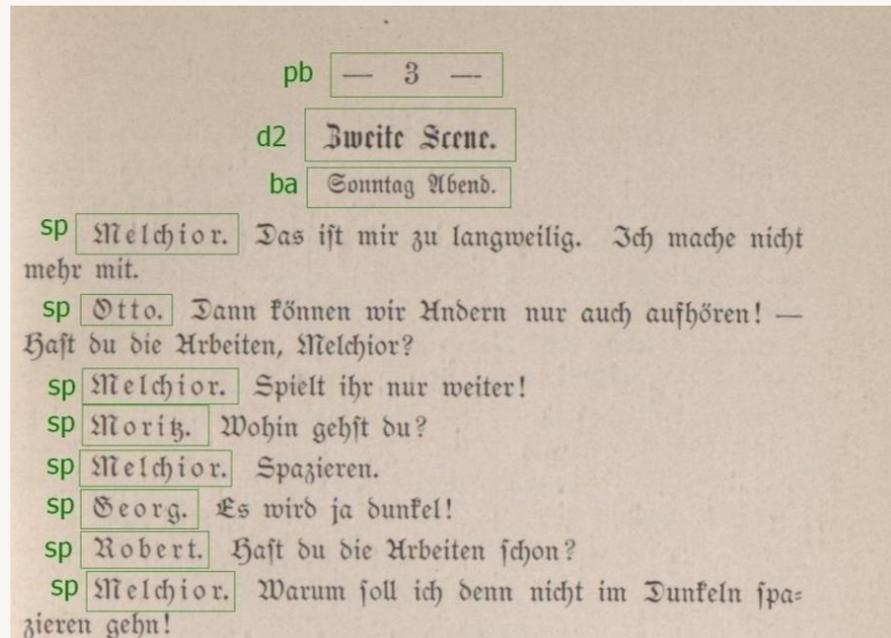
Quellen: Diagramm: Zusammensetzung der DTA-Korpora. Online: <https://www.deutschestextarchiv.de/doku/textauswahl>, abgerufen am 09.11.2021;  
Diagramm: Werke nach Digitalisierungsmethode. Online: <https://www.deutschestextarchiv.de/dtaq/stat/digimethod>, abgerufen am 09.11.2021

# Manuelle Erfassung

- Erfassung der Struktur des Textes
- Erfassung der Zeichen
  - zeichengenau nach Vorlage
  - keine Modernisierungen; keine Normalisierungen
- „Ich lasse mich nicht irre schreyn“
  - schreyn → schreyn 😊
  - schreyn → schreyn •
  - schreyn → schreien 😞



Quelle: Goethe, Johann Wolfgang von: Faust. Eine Tragödie. Tübingen, 1808, S. 287. In: Deutsches Textarchiv <[https://www.deutschestextarchiv.de/goethe\\_faust01\\_1808/293](https://www.deutschestextarchiv.de/goethe_faust01_1808/293)>, abgerufen am 09.11.2021.



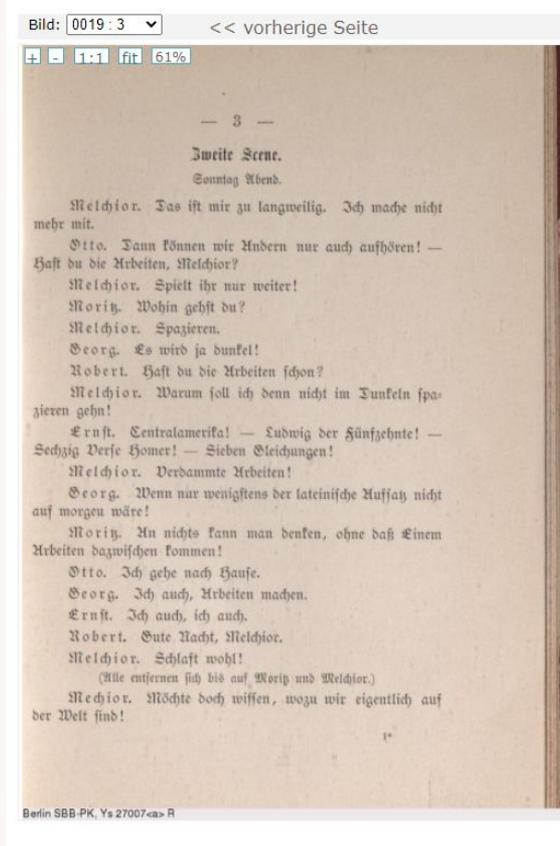
Quelle: Wedekind, Frank: Frühlings Erwachen. Zürich, 1891, S. 3. In: Deutsches Textarchiv <[https://www.deutschestextarchiv.de/wedekind\\_erwachen\\_1891/19](https://www.deutschestextarchiv.de/wedekind_erwachen_1891/19)>, abgerufen am 09.11.2021.

# Manuelle Erfassung

```
<pb="3"/>
<d2><b>Zweite Scene.</b></d2>
<ba>Sonntag Abend.</ba>
<sp><g>Melchior</g>.</sp> <p>Da&sr; ist mir zu langweilig. Ich mache nicht
mehr mit.</p>
<sp><g>Otto</g>.</sp> <p>Dann können wir Andern nur auch aufhören! &2014;
Hast du die Arbeiten, Melchior?</p>
<sp><g>Melchior</g>.</sp> <p>Spielt ihr nur weiter!</p>
<sp><g>Moritz</g>.</sp> <p>Wohin gehst du?</p>
<sp><g>Melchior</g>.</sp> <p>Spazieren.</p>
<sp><g>Georg</g>.</sp> <p>E&sr; wird ja dunkel!</p>
<sp><g>Robert</g>.</sp> <p>Hast du die Arbeiten schon?</p>
<sp><g>Melchior</g>.</sp> <p>Warum soll ich denn nicht im Dunkeln spa-
zieren gehn!</p>
```

Quelle: unveröffentlichte Rohtranskription, aus: Wedekind, Frank: Frühlings Erwachen. Zürich, 1891, S. 3. In: Deutsches Textarchiv  
<[https://www.deutschestextarchiv.de/wedekind\\_erwachen\\_1891/19](https://www.deutschestextarchiv.de/wedekind_erwachen_1891/19)>.

# Richtlinien zur TEI/XML-Annotation



```
<pb facs="#f0019" n="3"/>
<div n="2">
  <head> <hi rendition="#b">Zweite Scene.</hi>
</head><lb/>
  <stage> <hi rendition="#c">Sonntag Abend.</hi>
</stage><lb/>
  <sp who="#MEL">
    <speaker><hi rendition="#g">Melchior</hi>.
  </speaker>
  <p>Das i&#x017F;t mir zu langweilig. Ich
mache nicht<lb/>
mehr mit.</p>
  </sp><lb/>
```

Quelle: Wedekind, Frank: Frühlings Erwachen. Zürich, 1891, S. 3. In: Deutsches Textarchiv <[https://www.deutschestextarchiv.de/wedekind\\_erwachen\\_1891/19](https://www.deutschestextarchiv.de/wedekind_erwachen_1891/19)>, abgerufen am 09.11.2021.



# Zugänge zum Textarchiv

---

## Computer

- <https://www.deutschestextarchiv.de/api>

## Menschen

- DTA: <https://www.deutschestextarchiv.de/>
- DTA-Q: <https://www.deutschestextarchiv.de/dtaq>
- DTA-Basisformat: <https://www.deutschestextarchiv.de/doku/basisformat/>

... ich aber per quaestionum politicarum ambages, mit	<b>listigen</b>	<b>Fragen</b> (welches alle Medici in schweren Fällen ...
Eine	<b>nothwendige</b>	<b>Frage</b> ist es/ ob man des Abends ...
Job.	<b>12.</b>	<b>Frage</b> das Viehe/ das wird dichs lehren ...
... Offenbach, sah mich zweifelhaft an, that	<b>befremdende</b>	<b>Fragen</b> über mein Befinden, ich sah im ...
Ich bin überzeugt meine	<b>unbefangnen</b>	<b>Fragen</b> , um von ihm zu lernen, ...
Jetzt hab ich noch eine	<b>geringe</b>	<b>Frage</b> , aber sie gilt mir viel, ...
..., das ist deine Liebe zu mir:	<b>Geheimnißvolle</b>	<b>Frage</b> und unentbehrliche Antwort.
... Antworten/ Auff die im I. Theil vorgesetzte	<b>Allgemeine</b>	<b>Fragen</b> Von denen ketzereyen.
... ziemete dieses den Reformirten/ die in ihrer	<b>ersten</b>	<b>frag</b> bekäntniß thun/ daß ihr einiger trost ...

Quellen: Suchanfrage "\$p=ADJA Fragen", Online:

<https://www.deutschestextarchiv.de/search/ddc/search?ctx=8&q=%22%24p%3DADJA%20Fragen%22%20;start=1;limit=10;fmt=kwic>, abgerufen am 09.11.2021.

# t.evo – Die Evolution von komplexen Textmustern

Ein Projekt zur korpusbasierten Analyse  
von Textmusterwandel

# Übersicht zum Projekt

---

- Linguistische Erforschung von Textsortenmerkmalen bzw. Textmustern und deren Wandel
- DFG-finanziertes Projekt
- Laufzeit: 2019-2022
- Partner: Uni Paderborn – Lehrstuhl für Germanistische und Allgemeine Sprachwissenschaft; Zentrum Sprache der Berlin-Brandenburgische Akademie der Wissenschaften
- Leitung: Prof. Dr. Britt-Marie Schuster (Paderborn), PD Dr. Alexander Geyken (Berlin)

# Forschungsfrage und Projektziel I – Hintergrund

## Was ist Textlinguistik?

- Im Gegensatz zur Systemlinguistik (Syntax, Semantik, Morphologie und Phonetik/Phonologie) werden in der Textlinguistik größere sprachliche Einheiten als Wörter und Sätze untersucht
- Jene Einheiten hängen strukturell und funktional zusammen in einer Weise, die nicht (immer) anhand grammatischer Regeln erklärt werden kann
- Die Textlinguistik untersucht diese funktionalen und strukturellen Merkmale
  - Dazu zählen bspw.: Stilmittel, Thema, Sprachhandlungen usw.
- Textklassifikation ist ein weiteres Forschungsfeld der Textlinguistik
  - Der Begriff “Textmuster” meint hier strukturelle und funktionale Einheiten, die charakteristisch für eine Textsorte sein können (z. B. der Aufbau eines Kochrezepts)

# Forschungsfrage und Projektziel II

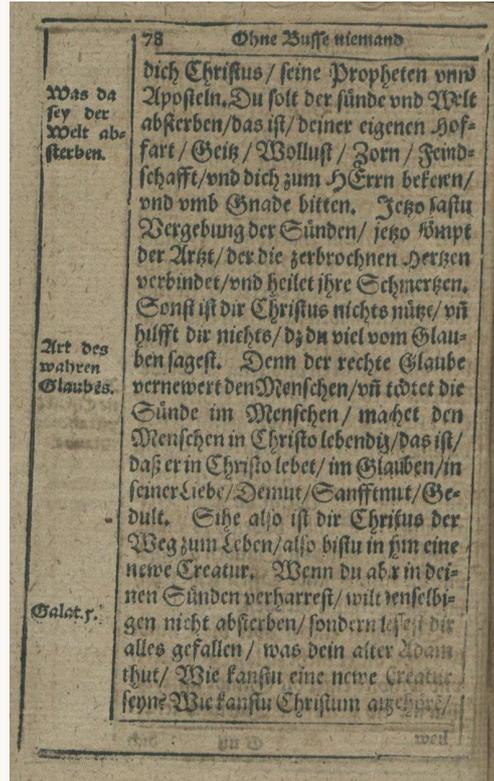
- Es werden zwei linguistische Forschungsbereiche zusammengeführt: Textlinguistik und diachrone Sprachforschung
- Typische Textmuster und deren Merkmale sollen ermittelt werden
  - dabei Rückgriff auf bisherige Forschungsergebnisse und textlinguistische Theorie, um Kategorien im Vorfeld festzulegen
- Anhand dieser Merkmale soll der Textmusterwandel beschrieben werden
- Operationalisierbarkeit und Anwendbarkeit mehrdimensionaler Textmodelle soll geprüft werden

# Datengrundlage I

## Welche Texte werden untersucht?

- 2 Textsorten der Gebrauchsliteratur
  - Presstexte des 19./20. Jhs. ((Augsburger) Allgemeine Zeitung) → 175 Texte
  - Erbauungsliteratur des 17./18. Jhs. (64 prosaische Erbauungsbücher ; 284 Funeralschriften) → 248 Texte

”Du solt der sünde vnd Welt absterben/ das ist/ deiner eigenen Hoffart/ Geitz/ Wollust/ Zorn/ Feindschafft/ vnd dich zum HERRN bekeren/ vnd vmb Gnade bitten. Jetzo sastu Vergebung der Sünden/ jetzo sömpt der Artzt/ der die zerbrochen Hertzen verbindet/ vnd heilet ihre Schmerzen. Sonst ist dir Christus nichts nütze/ vnd hilfft dir nichts/ dz dn viel vom Glauben sagest.”



SLUB Dresden, Theol.ev.asc.1269-1

Arndt, Johann: Von wahrem Christenthumb. Bd. 1. Magdeburg, 1610, S. 78.  
<[https://www.deutschestextarchiv.de/arndt\\_christentum01\\_1610/110](https://www.deutschestextarchiv.de/arndt_christentum01_1610/110)>



# Datengrundlage II

Kodierung der Struktur des Drucktextes  
(Absätze, Titel, Marginalien usw.)

TEI-XML

Kodierung des Wortartentaggings (bzw.  
sogenanntes POS-Tagging)

```
<div type="preface" n="2">
<head/>
<salute><hi rendition="#c"><hi rendition="#fr">Das walte die Heilige und
Hochge-<lb/> lobte Dreyeinigkeit/ Amen.</hi></hi></salute>
<lb/>
<p><hi rendition="#fr"><hi rendition="#in">G</hi>eliebte/ außerwählte</hi><lb/>
und zum theil auch &#x017F;chmertzlich Be-<lb/> tru&#x0364;bete/ Das
i&#x017F;t der Alten Krone/ wenn<note place="right"><hi rendition="#i"><hi
rendition="#aq">Sir. 25. v.</hi><lb/> 8.</hi></note><lb/>
&#x017F;ie viel erfahren haben/ und ihre Ehre<lb/> i&#x017F;t/ wenn
&#x017F;ie Gott fu&#x0364;rchten/ &#x017F;aget Sirach am <hi rendition="#aq">
XXV.</hi><lb/> Capitel. In welchen Worten der wei&#x017F;e Mann von
zwey-<lb/> erley Zierde alter ehrlicher Leute redet/ unter welchen<lb/> die
er&#x017F;te von ihm in &#x017F;einer Sprache genennet wird
&#x03D6;&#x03BF;&#x03BB;&#x03C5;&#x03D6;&#x03D6;&#x03B9;&#x03C1;&#x03AF;&#x03B1; viel erfahrung/
ver&#x017F;tehe vornemlich in &#x017F;einem Am-<lb/> pte und Beruff/ darein
Gott einen ge&#x017F;etzt/ welche erfah-<lb/> rung Sirach
heiß&#x017F;et;
&#x03C2;&#x03AD;&#x03C6;&#x03B1;&#x03BD;&#x03BF;&#x03BD;
&#x03B3;&#x03B5;&#x03C1;&#x03CC;&#x03BD;&#x03C4;&#x03C9;&#x03BD; eine Krone
der<lb/> Alten. Denn wie eine Krone das Haupt zieret eines<lb/>
Mens&#x017F;chen/ deme &#x017F;ie auffge&#x017F;etzt wird<hi rendition="#i">
></hi> al&#x017F;o auch die diel-<lb/> mannigfaltige Erfahrung der
Alten Leute/ bey welchen &#x017F;ie<lb/> gefunden wird. Die andere Zierde
alter Leute i&#x017F;t &#x03D5;&#x03CC;&#x03B2;&#x03BF;&#x03C2;<lb/>
&#x03BA;&#x03C5;&#x03C1;&#x03AF;&#x03BF;&#x03C5; des <hi rendition="#k">
>H&#x017F;rren</hi> fürchte/ da der Men&#x017F;ch Gott fu&#x0364;rch-<lb/> tet/
als ein liebes Kind &#x017F;einen fromen Vater/ und thut/ was<lb/>
dem&#x017F;elben gefa&#x0364;llig/ und hu&#x0364;tet &#x017F;ich mit allem
fleiß/ das es ihn<lb/> in keinerley wege oder wei&#x017F;e beleidige/ und
&#x017F;olche Gottes-<lb/> fürcht i&#x017F;t/ wie Sirach in &#x017F;einer
Sprache redet/ der Alten<lb/>
```

```
<salute><hi rendition="#c"><hi rendition="#fr"><s xml:id="s9"><w xml:id="wea"
lemma="die" pos="PDS" norm="Das">Das</w>
<w xml:id="web" lemma="walten" pos="VVPIN" norm="walte"
>walte</w>
<w xml:id="wec" lemma="d" pos="ART" norm="die">die</w>
<w xml:id="wed" lemma="heilig" pos="ADJA" norm="Heilige"
>Heilige</w>
<w xml:id="wee" lemma="und" pos="KON" norm="und">und</w>
<w xml:id="wef" lemma="hochgelobt" pos="ADJA" norm="Hochgelobte"
>Hochge-<lb/> lobte</w>
<w xml:id="wf0" join="right" lemma="Dreyeinigkeit" pos="NN"
norm="Dreyeinigkeit">Dreyeinigkeit</w><w xml:id="wf1"
join="left" lemma="/" pos="$ (" norm="/" /><lb/>
<w xml:id="wf2" join="right" lemma="Amen" pos="NN" norm="Amen"
>Amen</w><w xml:id="wf3" join="left" lemma="." pos="."
norm="." /></w></s></hi></hi></salute>
<lb/>
<p><hi rendition="#fr"><hi rendition="#in"><s next="#sa_2" xml:id="sa"><w
next="#wf4_1" xml:id="wf4" lemma="Geliebte" pos="NN"
norm="Geliebte">G</w></s></hi><s xml:id="sa_2" prev="#sa"
next="#sa_3"><w xml:id="wf4_1" prev="wf4" join="right"
>eliebte</w><w xml:id="wf5" join="left" lemma="/" pos="$ ("
norm="/" /></w>
<w xml:id="wf6" join="right" lemma="Außerwählte" pos="NE"
norm="außerwählte">außerwählte</w><w xml:id="wf7" join="left"
lemma="/" pos="$ (" norm="/" /></w></s></hi><lb/>
<s xml:id="sa_3" next="#sa_4" prev="#sa_2"><w xml:id="wf8" lemma="und"
pos="KON" norm="und">und</w>
<w xml:id="wf9" lemma="zum" pos="APPART" norm="zum">zum</w>
<w xml:id="wfa" lemma="Teil" pos="NN" norm="Teil">theil</w>
<w xml:id="wfb" lemma="auch" pos="ADV" norm="auch">auch</w>
<w xml:id="wfc" lemma="schmerzlich" pos="ADJD" norm="schmerzlich"
>fchmertzlich</w>
```

Quelle: Freitag, Christoph: Todes Trutz in Gottes Schutz. Oels, 1652. In: Deutsches Textarchiv <<https://www.deutschestextarchiv.de/537789>>

# Forschungsmethoden I – Überblick

- Mehrdimensionale Betrachtungsweise → Texte werden auf verschiedenen Ebenen (Stil, Thema, funktionale Textstrukturen wie bspw. Sprachhandlungen) analysiert
- Korpusbasierte Analyse → Es gibt eine Hypothese, die am Korpus geprüft werden soll, zu untersuchende Kategorien und Merkmale werden vorher festgelegt
- Automatisierte vs. manuelle Methoden
  - automatisierte quantitative Analyse mithilfe computerlinguistischer Methoden
  - manuelle Annotation nicht automatisch greifbarer Merkmale
  - Rückkopplung zwischen beiden
- Quantitative vs. qualitative Korpusanalyse
  - Quantitative Auswertung des Gesamtkorpus und der manuell annotierten Teilkorpora
  - Qualitative Analyse bei der manuellen Auswertung

# Forschungsmethoden II – Automatisierte Analyse

## Wie wird die automatisierte Analyse durchgeführt?

1. Auf Basis von Forschungsliteratur wird ein textlinguistisches Modell entwickelt. (in diesem Fall: mehrere Textebenen)
2. Das Modell wird “operationalisiert”, d. h. theoretische Kategorien werden in konkrete, empirisch messbare Kategorien übertragen, die sich computerlinguistisch untersuchen lassen (z. B. Reihung von Nomen, Kommunikationsverben).
3. Mithilfe von Pythonskripten wird der Text analysiert und die Ergebnisse ausgewertet.
4. Aus den statistischen Daten werden qualitative Rückschlüsse gezogen.
5. Bisherige Ergebnisse?

(Wiederholungsstrukturen→)

„**je mehr man** Gott siehet/ **je mehr man** jhn sehen will/  
**je mehr man** jhn liebet/ **je mehr man** jhn lieben will/  
vnd **je mehr man** jhn lobet/ **je mehr man** jhn loben will“

Arndt 1610a, Faks. 0319

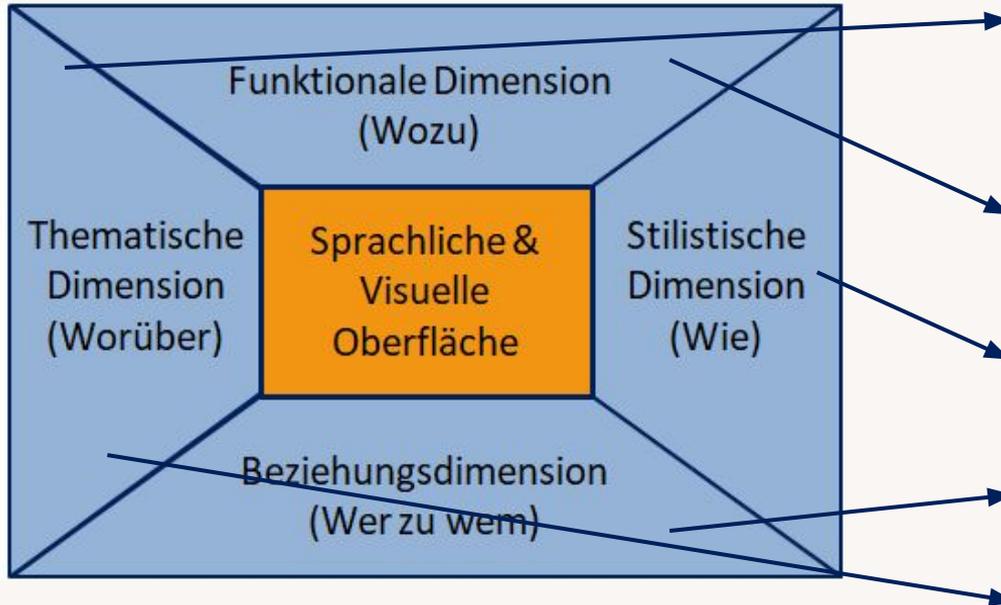


## Arbeitsschritte zur manuellen Annotation:

1. Auf Basis von Forschungsliteratur wird ein textlinguistisches Modell entwickelt (in diesem Fall: mehrere Textebenen).
2. Das theoretische Modell wird “operationalisiert”, d. h. theoretische Kategorien werden in konkrete, empirisch messbare Kategorien übertragen → in diesem Fall ein Tagset für die manuelle Annotation.
3. Ein Tagset wird erstellt, das digital angewandt werden kann und alle wichtigen Kategorien enthält.
4. Die manuelle Annotationspraxis mit diesem Tagset wird erprobt und angepasst
5. Das Korpus wird manuell annotiert.
6. Die Annotationen können quantitativ ausgewertet und qualitativ gedeutet werden.

# Forschungsmethoden III – Manuelle Annotation

## → Überführung des theoretischen Modells in konkrete Tags



Ebene 0: Formalia				
<ul style="list-style-type: none"> <li>Zitieren - direkt</li> <li>Zitieren - indirekt</li> </ul>		<ul style="list-style-type: none"> <li>Fremdsprachliches Material</li> <li>Übersetzen</li> </ul>		
Ebene 1: Quellspezifische Tags				
Predigt	Funeralschrift	Andachtsbücher	Gebet	Besondere Gestaltungsformen
<ul style="list-style-type: none"> <li>Einleitung</li> <li>Argumentatio</li> <li>Zurückweisung</li> <li>Explicatio</li> <li>Applicatio</li> <li>Epilog</li> </ul>	<ul style="list-style-type: none"> <li>Predigt</li> </ul>	<ul style="list-style-type: none"> <li>Schriftlesung</li> <li>Meditation</li> <li>Gebet</li> <li>Lied</li> </ul>	<ul style="list-style-type: none"> <li>Anrede Gottes</li> <li>Abbitte</li> <li>Bitte</li> <li>Fürbitte</li> <li>Dank</li> </ul>	<ul style="list-style-type: none"> <li>Potentielle Formeln</li> <li>Bibelbericht</li> </ul>
Ebene 2: Funktionale Textstrukturen (Sprachhandlungen)				
Direktiv-appellativ	Assertiv-informativ	Persuasiv	Konklusiv	Expressiv
<ul style="list-style-type: none"> <li>Anweisen-direkt</li> <li>Anweisen-indirekt</li> <li>Bitten</li> </ul>	<ul style="list-style-type: none"> <li>Auslegungskomplex</li> <li>Auslegen-Basis</li> <li>Auslegen-Bibel</li> <li>Auslegen</li> <li>Berichten</li> </ul>	<ul style="list-style-type: none"> <li>Behaupten-indirekt</li> <li>Eine-Meinung-äußern</li> </ul>	<ul style="list-style-type: none"> <li>Konklusion-Basis</li> <li>Erklären-warum</li> <li>Folgern</li> <li>Konzedieren</li> <li>Argumentieren-Bibel</li> <li>Zusammenfassen</li> </ul>	<ul style="list-style-type: none"> <li>Danken</li> <li>Klagen</li> <li>Lobpreisen</li> <li>Trösten</li> </ul>
Ebene 3: Übergreifende stilistische Handlungen				
<i>keine Unterkategorie</i>	Gewichten	Dialogisieren	Veranschaulichen	
<ul style="list-style-type: none"> <li>Bewerten</li> <li>Verfremden (Paradoxie)</li> </ul>	<ul style="list-style-type: none"> <li>Emphase</li> <li>Hervorheben</li> <li>Hervorheben-lyrisch</li> <li>Kontrastieren</li> <li>Übertreiben</li> </ul>	<ul style="list-style-type: none"> <li>Rhetorische Frage</li> <li>Fingierter Dialog</li> </ul>	<ul style="list-style-type: none"> <li>Anschaulich-belegen</li> <li>Beschreiben</li> <li>Exemplifizieren</li> <li>Illustrieren</li> <li>Vergleichen</li> </ul>	
Ebene 4: Beziehungskonstituierende sprachliche Handlungen				
Personen und Personengruppen		Metakommunikative Äußerung		
Textrezipient nennen	Fremdgruppe nennen	<i>keine Unterkategorie</i>		
<ul style="list-style-type: none"> <li>TR-inkludieren</li> <li>TR-exkludieren</li> </ul>	<ul style="list-style-type: none"> <li>Referenzperson</li> <li>Göttliche Entität</li> </ul>			
<i>keine Unterkategorie</i>				
<ul style="list-style-type: none"> <li>Textproduzent nennen</li> </ul>	<ul style="list-style-type: none"> <li>Quelle nennen</li> </ul>			
Ebene 5: Themen				
<ul style="list-style-type: none"> <li>Weltliche Ereignisse mit zeitgenössischem Bezug oder Bezug zur Lebenswelt</li> <li>Bezugnahme zur göttlichen Instanz oder zur Heilswelt (auch Leben und Wirken Christus)</li> <li>Folgerungen für das Christliche Leben</li> <li>Heilserwartung, eschatologische Dimension</li> </ul>				

Tagset der Erbauungsliteratur

# Forschungsmethoden III – Manuelle Annotation

## Arbeit in CATMA

- browserbasierte Anwendung zur Annotation
- Eigene Tagsets können erstellt werden
- Die Annotationsoberfläche ähnelt analogen Verfahren und Texteditoren
- Textabschnitt kann mehrere Annotationen gleichzeitig erhalten
- Statistische Analyse der Annotationen möglich

The screenshot displays the CATMA 6.3.1 web interface. The main window shows a document titled 'FREITAG TODESTRUTZ' with several lines of Latin text. Each line is annotated with colored bars (pink, blue, green, purple) indicating different categories. The text includes phrases like 'kurzte befehlen/ und ichreibt vom inhalt allo: David', 'in hoc Plal. à maximis periculis liberatus, & in regno', and 'Nach dem David auß großer Gefährlichkeit'. A sidebar on the left contains navigation options: 'Project', 'Tags', 'Annotate', and 'Analyze'. On the right, a panel titled 'Collection currently being edited' shows 'Freitag TodesTrutz Default Annotations' and a list of 'Tagsets' and 'Tags' with checkboxes for selection. The URL 'https://catma.de/' is visible at the bottom of the browser window.

<https://catma.de/>

# Forschungsmethoden III – Manuelle Annotation

## Welche Textmuster können sich in der manuellen Annotation zeigen?

*Beispiel: Auslegungskomplex*

**Magenta** = direktes Zitat (aus der Bibel)

**Blau** = Quellenverweis

**Graugrün** = Formel (Zitat)

**Grün** = Auslegungsbasis

**Rot** = Auslegung

Endlich wird der Sechste Vers dieses Pfalmens vom A-  
poftel zu den Ebræern am XIII. Capitel alfo wiederholet:  
Der Herr ift mein Helfer/ und wil mich  
nicht fürchten/ was folte mir ein Menich thun?

[Marginalie rechts: Ebr. 13, 6.]

welche worte aldar zu dem ende angezogen werden/ daß die  
Frommen fich nicht alleine verfürchern mögen das GOTT  
ihnen das geben werde/ was fie zu unterhaltung und Lebens-  
mitteln bedurfftig fein werden; fondern werden nach Da-  
vids Exempel allezeit an Gott einen folchen Helfer haben  
wieder ihre Feinde/ die neben dem Gutte auch des Bluttes  
der Frommen begehren. II. Auß vielen unfern Theolo-

Quelle: Freitag, Christoph: Todes Trutz in Gottes Schutz. Oels, 1652, S. [7].  
<<https://www.deutschestextarchiv.de/537789/7>>

# Projektstand und Ausblick

- **Stand von automatisierter Analyse und Annotation**
  - Die erste, automatisierte Analyse des gesamten Erbauungsliteraturkorpus ist abgeschlossen
  - Die Zeitungen für das Korpus der automatisierten Analyse befinden sich noch im Digitalisierungsprozess, die quantitative Analyse steht noch aus
  - Die manuelle Annotation der Erbauungsliteratur und Zeitungen steht noch aus, ebenso die quantitative Auswertung
- **Zusammenführung von manueller Annotation und automatisierter Analyse**
  - die manuelle Annotation soll die quantitative Analyse “flankieren”, indem sie Hinweise auf Merkmale und Strukturen liefert, die automatisiert analysiert werden können
- **Untersuchung des Textmusterwandels**
  - Das eigentliche Projektziel: Sobald Analyse aus beiden Methoden vorliegt und zusammengeführt wurde, können anhand der Ergebnisse qualitative Rückschlüsse auf den Wandel gezogen werden

Vielen Dank für Ihre  
Aufmerksamkeit!

Haben Sie noch Fragen?