

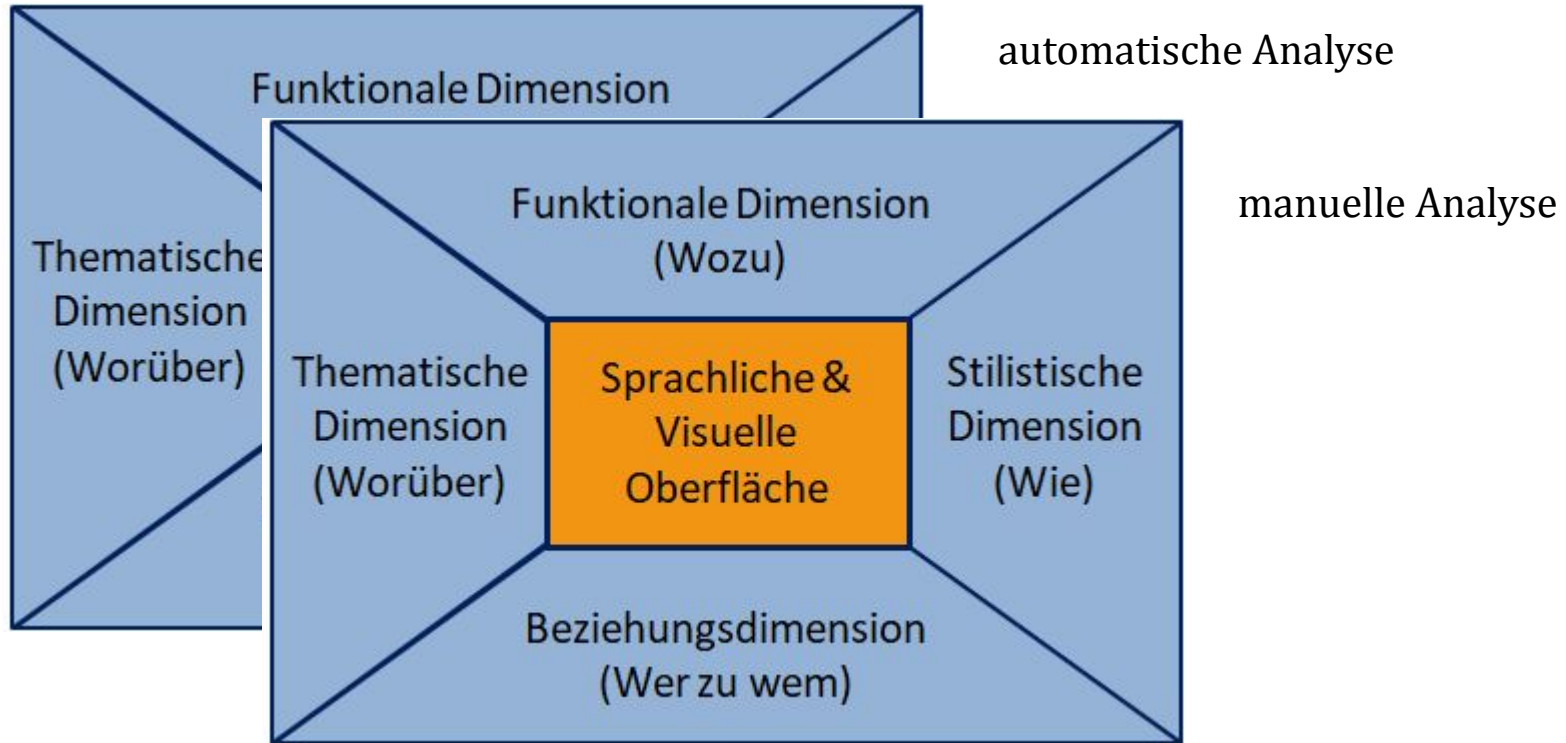
Mixed Methods in Aktion

Das Projekt t.evo und die mehrdimensionale Annäherung an das
Phänomen der Textsorte

Über das Projekt t.evo

- **Eckdaten**
 - 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)
- **Ziel**
 - typische Textmuster als Merkmale bestimmter Textsorten ermitteln
 - dabei Rückgriff auf bisherige Forschungsergebnisse
 - anhand dieser Merkmale Textmusterwandel beschreiben
- **Textsorten**
 - Presstexte des 19./20. Jhs. ((Augsburger) Allgemeine Zeitung)
 - Erbauungsliteratur des 17./18. Jhs. (prosaische Erbauungsbücher; Funeralschriften)
- **Verfahren**
 - manuelle Annotation & automatische Extraktion von Merkmalen
 - Kombination aus quantitativen und qualitativen Ansätzen
 - Mehrdimensionale Betrachtungsweise
 - Merkmale der untersuchten Textsorten & Generalisierbarkeit für andere Textsorten

Analysemodell in t.evo



Zeitung vs. Erbauungsliteratur → Äpfel und Birnen?

- **Gemeinsamkeiten:**
 - großes Publikum (Massen-Kommunikation)
 - Erbauungsliteratur war „bis in die zweite Hälfte des 18. Jh.s [...] für einen großen Teil der leidlich lese- und zuhörfähigen Bevölkerung als intensive Wiederholungslektüre die einzige oder vorwiegende Geistesbeschäftigung“ (von Polenz 2013: 23)
 - an die selektive Lektüre angepasstes Textdesign (vgl. Adamzik 2016: 280):
 - Presstexte: Cluster aus Fließtext, Grafiken, Bildern zum schnellen Überblick
 - Erbauliche Texte (als Sachliteratur): Fließtext, Bilder, Randbemerkungen, Register
 - Textsortennetz
 - zwei Genres, die jeweils ein Netz spezifischer Textsorten enthalten/hervorgebracht haben
- **Unterschiede:**
 - Thematik: profanes vs. geistliches Genre
 - Zeitschnitt: jeweils unterschiedlicher zeitlicher Zenit der Wirkmächtigkeit → für uns dadurch große diachrone Breite insgesamt
 - konkrete Gestaltungsmittel voraussichtlich einigermaßen divers → für uns große Breite an beobachtbaren Textmustern

Gliederung Vortrag

1. Digitalisierungsworkflow (Christopher Georgi)
2. Automatisierte Analyse (Susanne Haaf-Dumont)
3. Manuelle Annotation I – Pressetexte (Frauke Thielert)
4. Manuelle Annotation II – Erbauungsliteratur (Linda Kirsten)
5. Mixed Methods – Synthese beider Ansätze

Digitalisierungsworkflow

am Beispiel der Zeitungen

Überblick Digitalisierungsworkflow

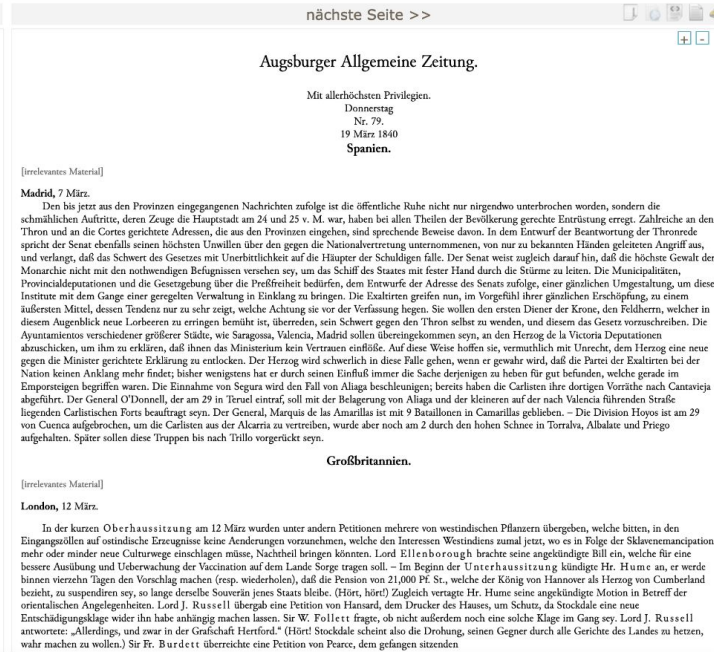
DTA

suchen Hilfe Texte ▼ Projekt ▼ Dokumentation ▼ Impressum
○ In den Titeldaten Im Korpus ○ In der Dokumentation

Allgemeine Zeitung, Nr. 79, Augsburg, 19. März 1840.



Vom Faksimile zum Digitalisat



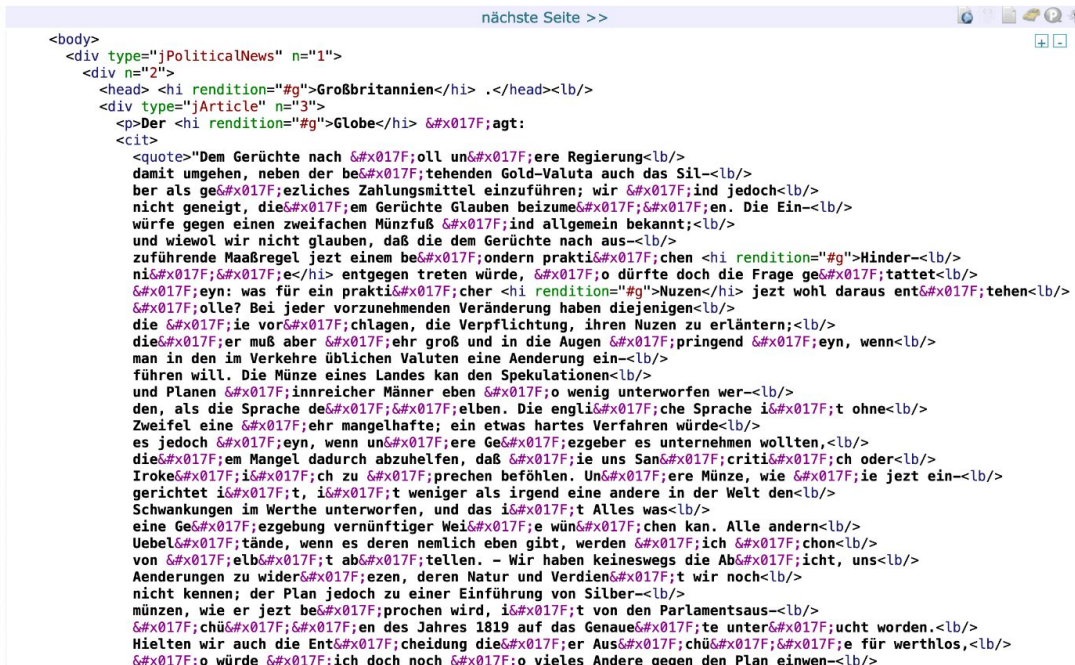
Überblick Digitalisierungsworkflow

Schritt 3: Konvertierung nach DTABF

nn_allgemeine01_1830 (t.evo)



offene Tickets: 0 (0 ganzes Buch) **Text** Text/Bild
 Stand: 2021-08-16 14:49:47 0-0-8 0-0-8



Überblick Digitalisierungsworkflow

Exkurs: Schwierigkeiten bei der Konvertierung – Strukturelle Besonderheiten



Entscheidung des Preisgerichts

Entwürfe für Hallenbad München-N

Erster Preis: Bildhauer Emil Wolf • Anstellung der Wettbewerbsteilnehmer

Das Preisgericht für die Entwürfe für ein Hallenbad in München-Nord hat am 12. Januar 1898 die Entwürfe von 12 Wettbewerbern beurteilt. Der erste Preis ist dem Bildhauer Emil Wolf zuerkannt worden. Die anderen Teilnehmer sind: Bildhauer Dr. Carl von Spreti, Architekt Dr. Carl von Spreti, Architekt Dr. Carl von Spreti, Architekt Dr. Carl von Spreti, Architekt Dr. Carl von Spreti, Architekt Dr. Carl von Spreti, Architekt Dr. Carl von Spreti, Architekt Dr. Carl von Spreti, Architekt Dr. Carl von Spreti, Architekt Dr. Carl von Spreti, Architekt Dr. Carl von Spreti, Architekt Dr. Carl von Spreti.

Die Reparaturen

Morgan und seine Hauptaufgabe

Der neue Wärtmeister hat das Wort • Erfolgreiche Befreiungen Gärten

Am 12. Januar 1898 hat der neue Wärtmeister des Schlosses in München sein Amt angetreten. Er hat sich sofort an die Hauptaufgabe gemacht, die Reparaturen der Gärten zu beenden. Er hat sich dabei auf die Hilfe der Gärtnerei verlassen, die sich in der letzten Zeit sehr erfolgreich bei der Befreiung der Gärten betätigt hat.

Gleicher Name, gleicher Geburtstag

Opfer eines seltsamen Gerichtsirrtums

Ein seltsamer Gerichtsirrtum hat in München zu einem traurigen Opfer geführt. Ein Mann, der am 12. Januar 1898 Geburtstag feierte, wurde fälschlicherweise für einen anderen Mann gehalten, der am selben Tag geboren wurde. Dies führte zu einer Verurteilung, die nach einer Klage durch den Angeklagten aufgehoben wurde.

Karneval

Am 12. Januar 1898 hat der Karneval in München seinen Höhepunkt erreicht. Die Teilnehmer haben sich in der Stadt herumgetrieben und haben viel Spaß gehabt. Die Organisatoren haben sich für die reibungslose Abwicklung des Festes bedankt.

1 Million Dollar

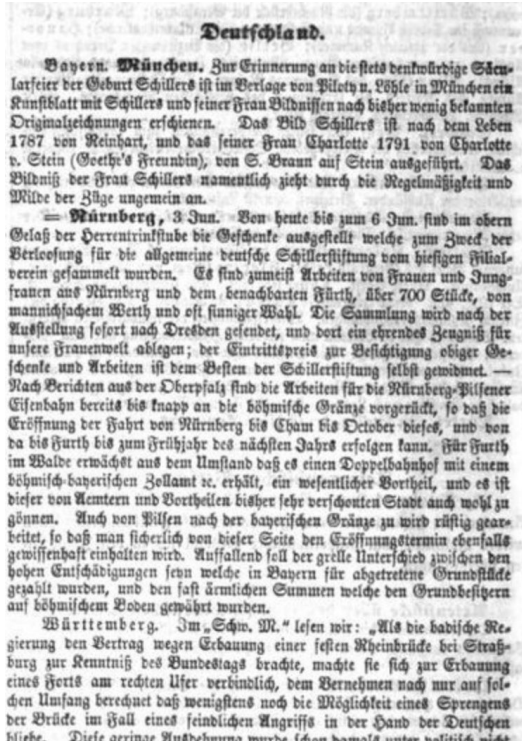
Sittung München für die Bekämpfung

Die Sittung München hat am 12. Januar 1898 beschlossen, eine Million Dollar für die Bekämpfung der Tuberkulose zu spenden. Dies ist ein großer Schritt in Richtung der Bekämpfung dieser Krankheit.

<div n="2" type="jArticle" xml:id="a01a" next="a01b">

Überblick Digitalisierungsworkflow

Exkurs: Schwierigkeiten bei der Konvertierung – Strukturelle Besonderheiten



```
<div type="jPoliticalNews"> <!-- Rubrik politische Nachrichten -->
  <div n="2"> <!-- Spezialgruppe politischer Nachrichten, z.B.
    "Deutschland" -->
    <head>[Subtitel, z.B. "Deutschland"</head>
    <div type="jArticle">[Politische Nachricht aus
      Deutschland]</div>
    <div type="jArticle">[Weitere politische Nachricht aus
      Deutschland]</div>
    </div>
    <div n="2"> <!-- Spezialgruppe politischer Nachrichten, z.B.
      "Frankreich" -->
      <head>[Subtitel, z.B. "Frankreich"</head>
      <div type="jArticle">[Politische Nachricht aus
        Frankreich]</div>
      <div type="jArticle">[Weitere politische Nachricht aus
        Frankreich]</div>
      </div>
    </div>
```

Überblick Digitalisierungsworkflow

Exkurs: Schwierigkeiten bei der Konvertierung – Abschreibfehler



Im Nebel

Zusammenstoß dreier Eisenbahnzüge
in den Vereinigten Staaten

Neuyork, 18. Januar. Bei Short Lane (Maryland) stießen gestern abend im Nebel drei Eisenbahnzüge zusammen. Zuerst waren zwei nach Norden fahrende Güterzüge ineinandergeraunt und dann fuhr ein nach Süden gehender Zug in die Trümmer mehrerer Güterwagen hinein. Vier Eisenbahnebeamte wurden getötet.

Überblick Digitalisierungsworkflow

Korpora:

Zeitungen

- 50 Ausgaben der Allgemeinen Zeitung zur manuellen Annotation (1830–1929)
- 125 weitere Ausgaben der Allg. Zeitung für quantitative Analysen

Erbauungsliteratur

- prosaische Erbauungsschriften & Gebetbücher: 64 Bände (17./18. Jh.), davon 37 Bände Eigendigitalisierungen
- Leichenpredigten: 284 Schriften (17. Jh.)

Unterscheidung Annotation und quantitative Analyse für beide Korpora

Automatisierte Analyse

am Beispiel der Erbauungsliteratur

Herangehensweise automatisierte Analyse

- **Datengrundlage**
 - TEI-Texte als Ergebnis Digitalisierung + DTA-Korpustexte
 - Volltexte nach automatisierter linguistischer Analyse (mit Lemma, POS, Satzgrenzen)
- **Merkmale im Korpusvergleich; hier:**
 - Korpus erbaulicher Prosaschriften (32 Bände, 17. Jh.)
 - Korpus von Leichenpredigten (284 Schriften, 17. Jh.)
 - Vergleichskorpus aus diversen Textsorten derselben Zeit (188 Bände, 17. Jh.)

Untersuchte Merkmale

Dimension	Funktional	Beziehung	Stilistisch	Thematisch	Verknüpfung
<u>Merkmal</u>					
<u>religiöser & mystischer Wortschatz</u>					
<u>mystische Wortbildung</u>					
<u>Sprechaktspezifischer Wortschatz</u>					
<u>Sprechaktspezifische Wortbildung</u>					
<u>Personaldeixis</u>					
<u>Feste Wendungen</u>					
<u>Reihungen</u>					
<u>Typische Satzarten</u>					
<u>Syntaktische Strukturen</u>					
<u>Negation</u>					
<u>Textkohärenz</u>					
<u>Textgliederung</u>					
<u>Textaufbau</u>					
<u>Intertextualität</u>					
<u>Rhetorische Figuren</u>					

- Merkmalsmatrix:
Dimension –
repräsentierende Merkmale
- im Folgenden zwei Beispiele:
 - Stil:
Wiederholungsstrukturen
 - Funktion/Thema:
Kommunikationsverben

Automatisiert und Quantitativ – die “5%-Hürde”?

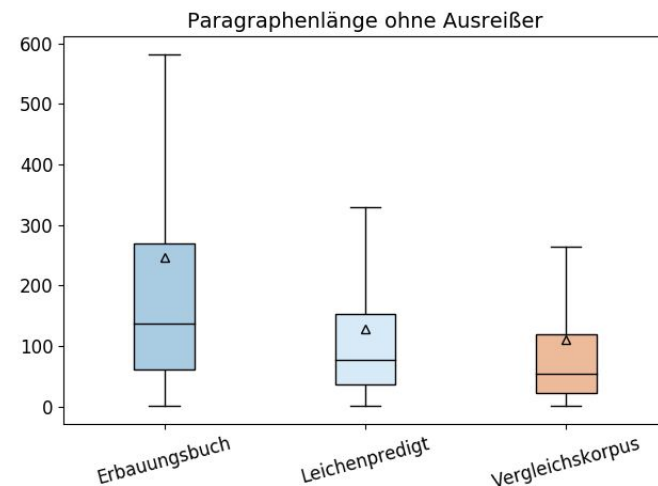
- **Statistische Analyse im Korpusvergleich**
 - Häufung, Verteilung und Streuung eines Merkmals in den betrachteten Korpora
 - nicht allein statistische Signifikanztests, sondern auch Evaluation der Ergebnisse der deskriptiven statistischen Analyse (z.B. Variationskoeffizient, Schiefe)
- **Vergleichende Analysen innerhalb eines Korpus**
 - Wie kommen Ausreißer zustande?
 - Welche Funktionen übernehmen Merkmale? Wie werden sie kompensiert, wo sie nicht auftreten?
- **Ergebnisse “auf Knopfdruck”?**
 - i.d.R. großer Aufwand bei der Vorverarbeitung & Merkmalsextraktion
 - für einige Merkmale nur teilautomatische Analyse möglich
 - Aufwand bei der Einordnung der Ergebnisse ([semi-]qualitativer Analyseschritt)

1. TEI-Tagging und n-Gramme: Wiederholungsstrukturen

- Erbauungsliteratur:
Tendenz zu Wiederholungen
- Operationalisierung hier auf
verschiedene Weisen
→ z.B. *Lemma-Bigramme je Paragraph*
 - Lemma: Grundform (*sagen* zu *sagte*)
 - Bigramm: Kombination zweier aufeinanderfolgender Lemmata (*je mehr | mehr man | man Gott | Gott siehet*)
 - Paragraph: TEI-Tagging fließt in die Auswertung ein
- Problem
 - Je länger ein Paragraph, desto wahrscheinlicher gibt es Wiederholungen.
 - daher: Berücksichtigung der durchschnittlichen Paragraphenlängen

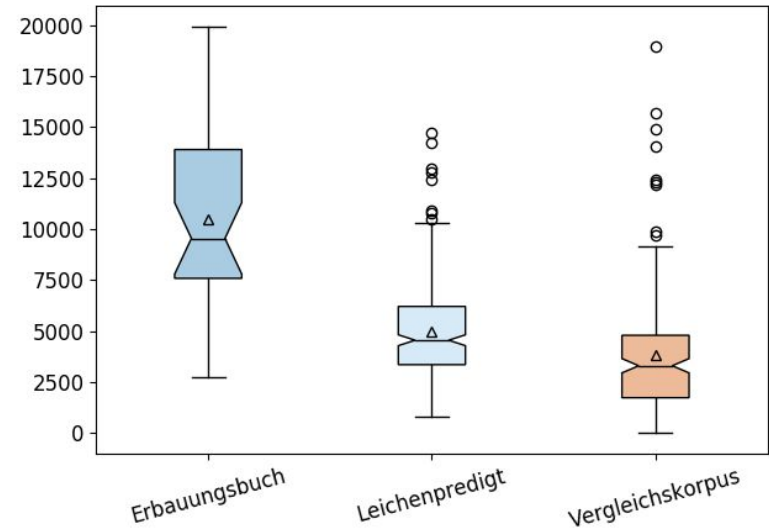
„**je mehr man** Gott siehet/ **je mehr man** jhn sehen will/
je mehr man jhn liebet/ **je mehr man** jhn lieben will/
vnd **je mehr man** jhn lobet/ **je mehr man** jhn loben will“

Arndt 1610a, Faks. 0319



1. TEI-Tagging und n-Gramme: Wiederholungsstrukturen

- Ergebnisse Lemma-Bigramm-Wiederholungen in Paragraphen
 - sehr relevant in erbaulichen Prosaschriften
 - weniger relevant in Leichenpredigten
 - ergo: Relevanz abhängig von der erbaulichen Textsorte
- Funktion des Merkmals
 - Mittel der Verständnissicherung (Erbauungsliteratur will Weisung geben zum rechten Glauben und christlichen Leben)
 - Emotionalisierende Wirkung (EL will mitreißen [movere])
 - Wiederholung (ruminatio) ist ein wichtiger Bestandteil der Meditation (EL will zur Meditation anregen)



2. Lexik: Analyse von Kommunikationsverben

- Kommunikationsverben, um sprachliche Handlungen zu vollziehen
 - Bsp.: “Sie gingen mit ihren Laternen hinaus und sangen.” → Sprechen über eine Handlung
 - Bsp.: “Ich *schlage* vor, wir machen das gemeinsam.” → Handeln mithilfe von Sprache
 - Bsp.: “Ich *taufe* Dich im Namen des Vaters, des Sohnes und des Heiligen Geistes.”
- Probleme:
 - 1. Nicht jede sprachliche Handlung benötigt auch ein Kommunikationsverb
 - Bsp.: “Bitte schließe das Fenster.”
 - 2. Je nach Kontext kann dasselbe Kommunikationsverb ganz unterschiedliche sprachliche Handlungen mit stützen
 - Bsp.: “Er *schrieb* ihr einen Brief.” vs. “Sie wollte ihm *schreiben*, aber es kam immer etwas dazwischen.”

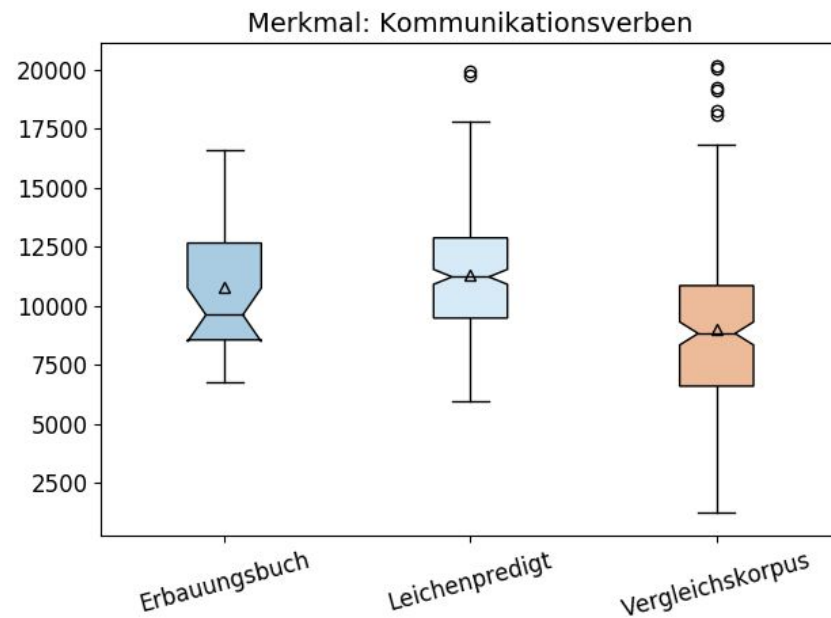
2. Lexik: Analyse von Kommunikationsverben

- Was können wir mithilfe von Kommunikationsverben ermitteln?
 - Eher nicht: Sprachhandlungen in den untersuchten Texten
 - Dafür aber: Thematisierung kommunikativer Akte → Häufung des Redens über Kommuniziertes oder zu Kommunizierendes (im Text, durch Dritte, zwischen Rezipient und Gott, etc.)
- Herangehensweise hier:
 - Quantitative Erhebung von Kommunikationsverben in verschiedenen Textsorten
 - Qualitative Auswertung unter Zuhilfenahme von Kollokationsanalysen mit DiaCollo (Jurish 2015)

2. Lexik: Analyse von Kommunikationsverben

Ergebnisse:

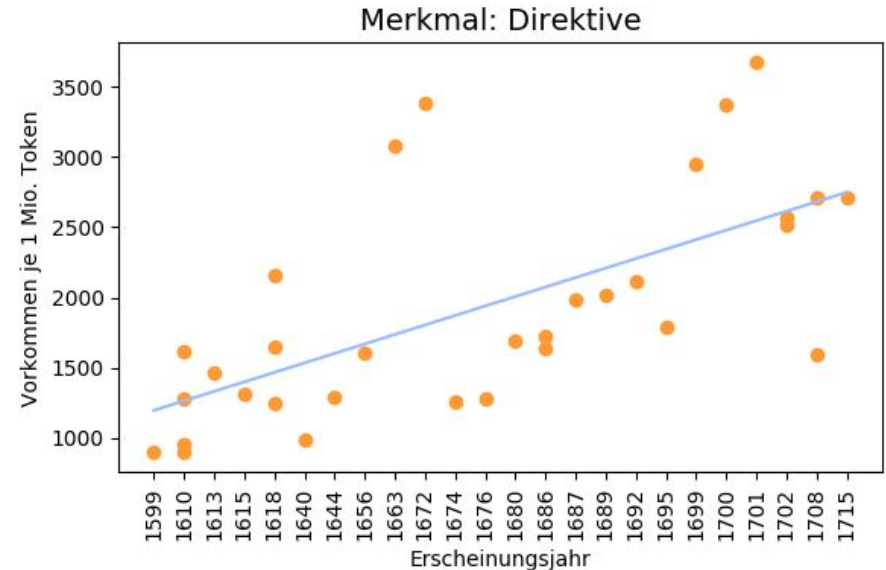
- Kommunikationsverben in erbaulichen Textsorten
 - stärkere Häufung
 - homogener verteilt... als in anderen Textsorten derselben Zeit
- herausragend mit deutlichem Abstand:
 - Verba dicendi allgemein (5 Verben, darunter: **sagen**, sprechen, reden)
- *sagen* dabei sehr häufig Hinweis auf Zitate/intertextuelle Bezugnahmen:
 - **Paulus** *faget* zu den Eltesten zu Epheso
 - Drumb *faget* **David** im 34. Pfalm. v. 20
 - was **Chriftus** *faget*/ Luc. 9. v. 23



2. Lexik: Analyse von Kommunikationsverben

Ergebnisse:

- in prosaischen Erbauungsbüchern nehmen die Direktiven im zeitlichen Verlauf deutlich zu
 - fragen, befehlen, fordern
- wichtig für die erbaulichen Textsorten sind außerdem die Expressiven
 - Ausdruck von Emotionen
 - z.B. besonders Lütkemann: danken, loben, preisen



Ergebnisse

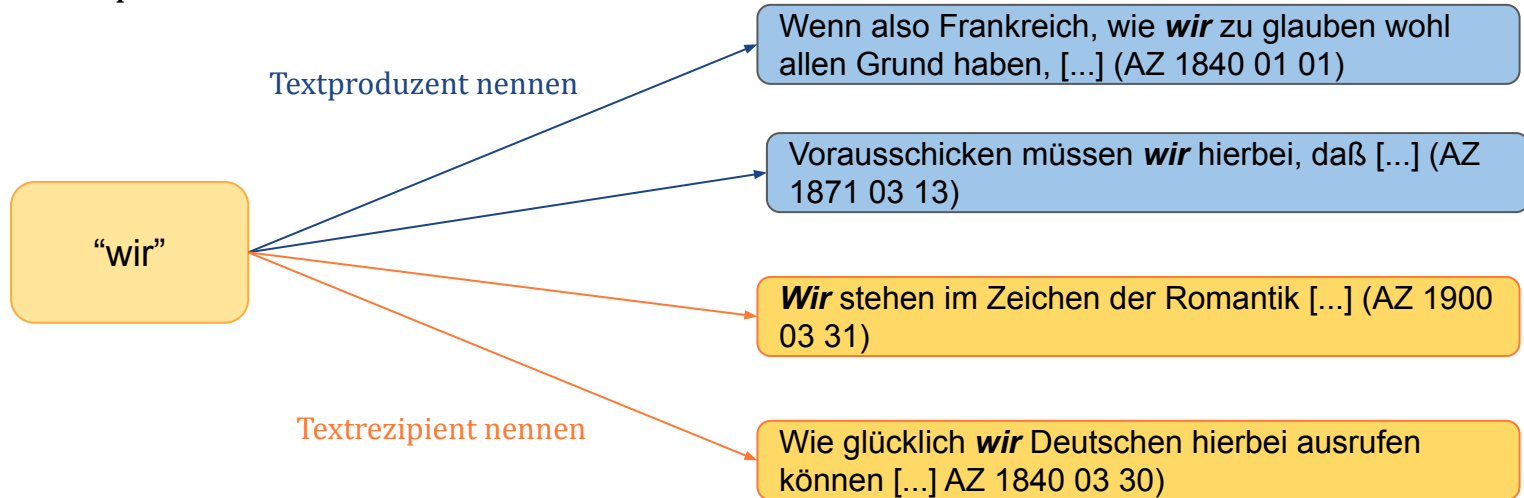
Analyse-dimension	Textebene	Hilfsmittel Operationalisierung	Funktion	Umsetzung
Stil	Syntagma	TEI-Tagging + Lemmatisierung	Verständnissicherung Anrühren → Überzeugen Anregung zur Meditation	Wiederholungen
Funktion/ Thema	Lexik	Lemmatisierung + POS-Tagging	Rezipienten für die Sache einnehmen: <ul style="list-style-type: none">• Emotionen ansprechen• Inhalte an Autoritäten rückbinden	Verwendung bestimmter (Typen von) Kommunikationsverben

Manuelle Annotation I

am Beispiel der Pressetexte

Manuelle Annotation - Generelles

- Annotation linguistischer Phänomene in allen Dimensionen des Analysemodells, die durch automatisierte (noch) nicht erfasst werden können
 - Bsp.: das Pronomen “wir”



Manuelle Annotation - Tagset

- Manuelle Annotation erfasst nur zuvor definierte sprachliche Einheiten→ mehrschrittiges Verfahren
 - Schritt 1: Festlegung der zu annotierenden sprachlichen Einheiten anhand der aktuellen Forschungsliteratur
 - Schritt 2: Überprüfung der Plausibilität des Tagsets anhand Annotation eines Pilotkorpus
 - Schritt 3: Anpassung des Tagsets an die Textrealitäten
 - Schritt 4: Erstellung von Guidelines zur Dokumentation und Qualitätssicherung
- Aufbau und Struktur
 - sprachhandlungsorientiert (vom Textproduzent ausgehend die Sprachhandlung benennend)
 - Tagset bestehend aus 6 Ebenen:
 - bilden die Dimensionen des Analysemodells ab (Ebenen 3-5)
 - Ergänzung zweier Ebenen, um den Besonderheiten der Pressekommunikation(bzw. der Erbauungsliteratur) gerecht werden zu können (0-1)
 - Verzahnung der Ebenen:
 - zweiteiliges Tagset bestehend aus groß- und kleinräumigen Tags
 - Gesamtartikel umfassende Annotation von Thema und Textsorte→ analysierbare Einheiten innerhalb derer textmusterrelevante Phänomene erhoben werden können

Manuelle Annotation - Tagset

Ebene 0: Formalia und Layout			
Formalia		Layout	
<ul style="list-style-type: none"> • Zitieren - direkt • Zitieren - indirekt • Fremdsprachliches Material • Übersetzen 		<ul style="list-style-type: none"> • Überschrift • Zwischenüberschrift • Rubrik 	
Ebene 1: Quellenspezifische Tags - Pressekommunikation			
Textsorten		Gestaltungsformen	
<ul style="list-style-type: none"> • Kommunikat • Meldung • Nachricht • Verlaufsbericht 		<ul style="list-style-type: none"> • Hintergrundbericht • Rationierender Bericht • Kritik oder Rezension • Feuilleton 	<ul style="list-style-type: none"> • Anmerkung (Red.) • Korrespondenzblock • Epistemische Äußerung • Potentielle Formel
Ebene 2: Charakteristische Sprachhandlungen			
Informierende Sprachhandlungen		konklusive Sprachhandlungskomplexe	
Berichten über <ul style="list-style-type: none"> • Ereignis • Zustand • Ergebnis 	Beschreiben <ul style="list-style-type: none"> • Person • Ort • Gegenstand 	<ul style="list-style-type: none"> • KS-gesamt • KS-Basis • KS-erklären • KS-begründen • KS-rechtfertigen • KS-etwas folgern / schließen • KS-etwas zusammenfassen 	
Das Berichtete <ul style="list-style-type: none"> • einordnen • darstellen 		Einen Bericht <ul style="list-style-type: none"> • fortsetzen • wiedergeben • verweisen auf 	
persuasive Sprachhandlungen		Perspektivierende Sprachhandlungen	
Eine Meinung <ul style="list-style-type: none"> • wiedergeben • äußern 	Einen Einwand <ul style="list-style-type: none"> • anbringen • zurückweisen 	<ul style="list-style-type: none"> • Abwägen • Spekulieren • Wünschen/Hoffnung äußern 	
Ebene 3: Stil			
Bewerten <ul style="list-style-type: none"> • ist gleich • ist anders 	Gewichten <ul style="list-style-type: none"> • Emphase • Kontrastieren • Hervorheben 	Dialogisieren <ul style="list-style-type: none"> • rhet. Frage • fing. Dialog • Aufforderung • concessio 	Veranschaulichen <ul style="list-style-type: none"> • Beschreiben / erzählen • Illustrieren
Ebene 4: Beziehungsdimension			
Personen und Personengruppen		Raum und Zeit	
Informant nennen <ul style="list-style-type: none"> • I-benannt • I-unbenannt 	Ohne Differenzierung <ul style="list-style-type: none"> • Textproduzent nennen • Textrezipient nennen • Fremdgruppe nennen 	Handlungsraum <ul style="list-style-type: none"> • deiktisch • andere 	Handlungszeit <ul style="list-style-type: none"> • deiktisch • andere
Verständnissicherung		Ohne Differenzierung	
<ul style="list-style-type: none"> • Metakommunikative Äußerung 		<ul style="list-style-type: none"> • Berichtsort 	<ul style="list-style-type: none"> • Berichtszeit
Ebene 5: Thematische Dimension			
Innenpolitik <ul style="list-style-type: none"> • IP-Inland • IP-Ausland 	Außenpolitik <ul style="list-style-type: none"> • AP-Inland • AP-Ausland 	Wirtschaft/Börse <ul style="list-style-type: none"> • WB-Inland • WB-Ausland 	Ohne Differenzierung <ul style="list-style-type: none"> • Lokales • Aus aller Welt • Gesellschaft • Religion/Kirche • Wissenschaft und Kultur • Sport

Ebene 0 - Formales/Layout: formale Besonderheiten der Textsorte Pressekommunikation (7 Tags)

Ebene 1- Quellenspezifische Tags: Textsorten der Pressekommunikation & pressenspezifische Gestaltungsmittel (12 Tags)

Ebene 2 - charakteristische Sprachhandlungen (funktionale Dimension): vier Kategorien entsprechend den Funktionen der Pressekommunikation (25 Tags)

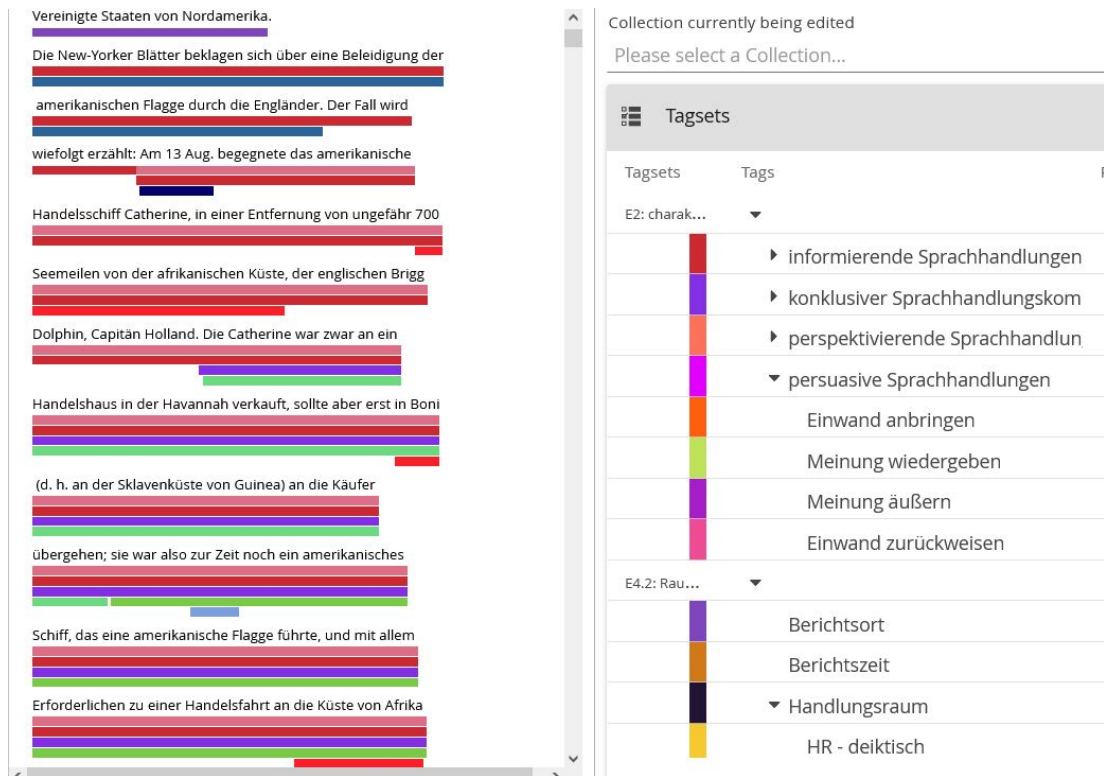
Ebene 3 - Stil (stilistische Dimension): vier Kategorien stilistischer Sprachhandlungen (11 Tags)

Ebene 4 - Beziehungsdimension: Personen und Personengruppen, Raum und Zeit & Maßnahmen der Verständnissicherung (12 Tags)

Ebene 5 - Thematische Dimension: "journalistische Ressorts" (12 Tags)

Manuelle Annotation - Catma 6.3.1

- Vorteile:
 - Erstellung von taxonomisch aufgebauten individuellen Tagsets beliebiger Komplexität
 - Anpassungen/Ergänzungen des Tagsets jederzeit möglich
 - Analysefunktion
- Nachteile:
 - nicht für große Textmengen geeignet
 - Visualisierung von TEI-Strukturen nicht vorgesehen
 - keine Ansicht von Originaldokumenten



Vereinigte Staaten von Nordamerika.
Die New-Yorker Blätter beklagen sich über eine Beleidigung der
amerikanischen Flagge durch die Engländer. Der Fall wird
wiefolgt erzählt: Am 13 Aug. begegnete das amerikanische
Handelsschiff Catherine, in einer Entfernung von ungefähr 700
Seemeilen von der afrikanischen Küste, der englischen Brigg
Dolphin, Capitän Holland. Die Catherine war zwar an ein
Handelshaus in der Havannah verkauft, sollte aber erst in Boni
(d. h. an der Sklavenküste von Guinea) an die Käufer
übergehen; sie war also zur Zeit noch ein amerikanisches
Schiff, das eine amerikanische Flagge führte, und mit allem
Erforderlichen zu einer Handelsfahrt an die Küste von Afrika

Collection currently being edited
Please select a Collection...

Tagsets

Tagsets	Tags
E2: charak...	<ul style="list-style-type: none">informierende Sprachhandlungenkonklusiver Sprachhandlungskomperspektivierende Sprachhandlungpersuasive SprachhandlungenEinwand anbringenMeinung wiedergebenMeinung äußernEinwand zurückweisen
E4.2: Rau...	<ul style="list-style-type: none">BerichtsortBerichtszeitHandlungsraumHR - deiktisch

Mixed Methods in Aktion - Potentielle Formeln

- Als Potentielle Formeln werden in der manuellen Annotation sprachliche Einheiten erfasst, die für Presstexte typische Funktionen übernehmen, aber nicht der Darstellung des Inhalts eines Artikels dienen, z.B. :
 - Quellenangaben
 - Kennzeichnung von Text- und Redewiedergabe
 - Thematisierung von Kommunikationsstrategien
 - Spekulation über vergangene und zukünftige Ereignisse
 - ...
- Hypothese: standardisierte Formulierung entstehen vor allem dort, wo ein Sprachhandlungstyp, ein Aussagegehalt oder ein Themenbezug häufig wiederholt werden müssen, ohne dass dabei stilistische Normen oder Performativität eine Rolle spielen
- Ziel: Ergebnisse der manuellen Annotation für Erhebung Potentieller Formeln durch automatisierte Verfahren nutzen

Mixed Methods in Aktion - Potentielle Formeln

- Vorgehen:
 - Manuelle Annotation aller sprachlicher Einheiten, die mit dem Ausdruck der genannten Textsortenfunktionen des Mediums Zeitung zusammenhängen
 - Auswertung über die Analysefunktion in Catma (KWIC-Ansicht)

Document	Left Context	Keyword	Right Context	Sta
AZ 1840 01 ...	AZ Finanzen von Grund ...	(Hannover'sche Ztg.)	Oesterreich. + *Wien	45
AZ 1900 04 ...	AZ Chabarowsk, 1. April....	(Meldung der R[...]hen-Agentur.)	Nach erfolgter Prüfu...	41
AZ 1900 04 ...	AZ Ferdinand Schöning...	Alein über da[...] zu erheben,	denn lichere Anzeich...	39
AZ 1840 01 ...	AZ Fünftel des gesamt...	Angefangen sin[...]eiten schon:	von Ans bis Lüttich auf	65
AZ 1900 03 ...	AZ . K. C.	Aus Bombay wir[...] berichtet:	In Pefchauer hat lich ...	41
AZ 1871 03 ...	AZ Grant Luft zu machen.	Das „ Siècle “[...]t Entrüstung:	„Kein Wort der Sym	50
AZ 1900 03 ...	AZ meiften deutlichen St...	Dem Vernehmen nach	haben die Militärver...	30
AZ 1840 01 ...	AZ Flagge durch die Eng...	Der Fall wird [...]olgt erzählt:	Am 13 Aug. begegnete	23
AZ 1900 04 ...	AZ Hochschule in Bresla...	Der Kultusminifter erklärt,	diefe Idee fei ihm fym...	84

Mixed Methods in Aktion - Potentielle Formeln

- Ermittlung musterhafter Lexemverbindungen/musterhafter Merkmale
 - Auswertungsbeispiel: Angabe der Quelle

Form	Realisierung
„nach“/„laut“ + Quelle	Nach Berichten von xy, nach Briefen aus xy,
Präp + Berichtsort + Passivkonstruktion	Aus xy wird geschrieben, aus xy wird gemeldet
Unpersönliche Ausdrücke + verba dicendi et sentiendi	Es heißt, es wird berichtet, es wird gemeldet, man behauptet
wie + (Pronomen) + verba dicendi et sentiendi	Wie man hört, wie gesagt, wie wir hören, wie es heißt
Informationstyp / Informationsgewinnung + „zufolge“ / „nach“	Berichten/Reuter zufolge, dem Vernehmen/ Anschein nach
Quellennennung in Klammerung	(National), (Times), (Reg. Zeitg.)

Manuelle Annotation II

Erbauungsliteratur

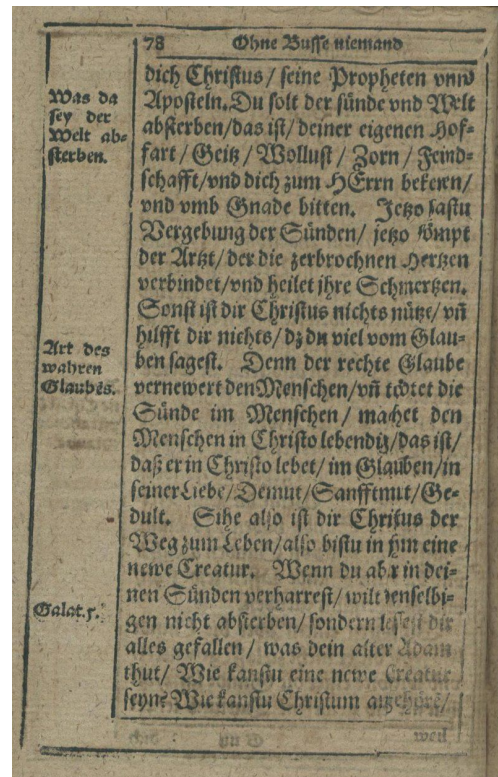
Manuelle Annotation des Erbauungsliteraturkorpus

Erbauungsliteraturkorpus für die manuelle Annotation:

- Texte aus dem 17./18. Jahrhundert
- insgesamt 50 Texte mit jeweils 20–30-seitigen Auszügen
- Leichenpredigten, Andachtsbücher und Gebetbücher

Auf folgende Aspekte soll hier eingegangen werden:

- Anpassung des Tagsets für die Erbauungsliteratur
 - Welche Kategorien müssen ergänzt / gestrichen werden?
- Beispiele für Ergänzung der automatisierten Analyse durch manuelle Annotation
 - Charakteristische Sprachhandlungen
 - Nennung des Textproduzenten



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

https://www.deutschestextarchiv.de/arndt_christentum01_1610/110

Anpassung des bestehenden Tagsets

Ebene 0: Formalia und Layout			
Formalia		Layout	
Zitieren - direkt		Überschrift	
Zitieren - indirekt		Zwischenüberschrift	
Fremdsprachliches Material		Rubrik	
Übersetzen			
Ebene 1: Quellspezifische Tags - Pressekommunikation			
Textsorten		Gestaltungsformen	
Kommunikat	Hintergrundbericht	Anmerkung (Red.)	Epistemische Äußerung
Meldung	Räsonierender Bericht	Korrespondenzblock	
Nachricht	Kritik oder Rezension		Potentielle Formel
Verlaufsbericht	Feuilleton		
Ebene 2: Charakteristische Sprachhandlungen			
Informierende Sprachhandlungen		Konklusive Sprachhandlungskomplexe	
Berichten über	Beschreiben	KS-gesamt	
Ereignis	Person	KS-Basis	
Zustand	Ort	KS-erklären	
Ergebnis	Gegenstand	KS-begründen	
Das Berichtete	Einen Bericht	KS-rechtfertigen	
einordnen	fortsetzen	KS-etwas folgern / schließen	
darstellen	wiedergeben	KS-etwas zusammenfassen	
	verweisen auf		
persuasive Sprachhandlungen		Perspektivierende Sprachhandlungen	
Eine Meinung	Einen Einwand	Abwägen	
wiedergeben	anbringen	Spekulieren	
äußern	zurückweisen	Wünschen / Hoffnung äußern	
Ebene 3: Übergreifende stilistische Handlungen			
Bewerten	Gewichten	Dialogisieren	Veranschaulichen
ist gleich	Emphase herstellen	rhet. Frage	Beschreiben / erzählen
ist anders	Kontrastieren	fing. Dialog	Illustrieren
	Hervorheben	concessio	
		Aufforderung	
Ebene 4: Beziehungskonstituierende sprachliche Handlungen			
Personen und Personengruppen		Raum und Zeit	
Informant nennen	Ohne Differenzierung	Handlungsraum nennen	Handlungszeit nennen
I-benannt	Textproduzent nennen	deiktisch	andere
I-unbenannt	Textrezipient nennen	andere	andere
	Fremdgruppe nennen		
Verständnissicherung		Ohne Differenzierung	
Metakommunikative Äußerung		Berichtsort nennen	Berichtszeit nennen
Ebene 5: Thema			
Innenpolitik	Außenpolitik	Wirtschaft/Börse	Ohne Differenzierung
IP-Inland	AP-Inland	WB-Inland	Lokales
IP-Ausland	AP-Ausland	WB-Ausland	Aus aller Welt
			Gesellschaft
			Religion/Kirche
			Wissenschaft und Kultur
			Sport

Layout ist weniger komplex,
Oberkategorie wird entfernt

andere quellspezifische Kategorien

Oberkategorien werden z. T. ergänzt
bzw. entfernt

Unterkategorien werden entfernt bzw.
quellspezifisch ergänzt

nicht relevant, wird entfernt

andere quellspezifische Themen

Tabelle 1: Tagset für die Annotation des Teilkorpus „Manuelle Annotation A2“ in Catma 6.3.0.

Rückkopplung manueller und automatisierter Methoden

Wie kann die manuelle Annotation textlinguistische Merkmale der Erbauungsliteratur erfassen, die mit der automatisierten Analyse nicht greifbar waren?

Kleine Analyse anhand folgender Merkmale:

- Beispiel I: Sprachhandlungen
- Beispiel II: Nennung des Textproduzenten

Beispiel I: Charakteristische Sprachhandlungen

- Handlungen, die mit Sprache vollzogen werden können, sind bspw.: *eine Meinung äußern, jmdn. um etwas bitten*
- Diese Handlungen können explizit mithilfe eines entsprechenden Verbs signalisiert werden ODER ergeben sich aus dem inhaltlichen/sprachlichen/sozialen Kontext
 - “Ich bitte dich, mir das Glas zu reichen.” → Das Verb “bitten” signalisiert die Sprachhandlung “Bitten”
 - “Es wäre schön, wenn du mir das Glas reichen könntest.” → Die Sprachhandlung “Bitten” ist implizit
 - “Er bittet dich, ihm das Glas zu reichen.” → verschiedene Lesarten, “bitten” signalisiert nicht zwangsläufig die Sprachhandlung “Bitten”
- In der automatisierten Analyse hat sich gezeigt: Kommunikationsverben verweisen nicht eindeutig auf Sprachhandlungen
- Hypothese: Sprachhandlungen in der Erbauungsliteratur sind mehrheitlich nicht explizit anhand von Kommunikationsverben gekennzeichnet.

Beispiel I: Charakteristische Sprachhandlungen

[...] daß wir klagen müssen mit König David: *Vnsere Seele sey betrübet/ vnd vnruhig in vns/ auß dem 42. Psalm.*

Solche Vnruh richtet vns Satanas zu/ der vnsaubere/ vnruhige Geist/ Luc. 11. der das Land durchzeucht/ Hiob 1. Ob er nemblich jemand möchte berücken/ vnd vnter seine Gewalt bringen. Der ist der brüllende Löwe/ der da suchet/ welchen er verschlinge/ in der 1. Epistel Petri 5. cap. [...]

<https://www.deutschestextarchiv.de/359997/19>

Inhaltlicher Rückbezug

- die Sprachhandlung wird durch den inhaltlichen Rückbezug vollzogen, nicht bestimmte Verben

Verba dicendi

- Im vorliegenden Beispiel wird damit kein assertiver Sprechakt vollzogen, sondern nur beschrieben

Der Apostel Petrus saget in seiner ersten Epistel cap. 4. vnd spricht: *Welche da leiden nach GOTTes Willen/ die sollen Jhm jhre Seelen befehlen/ als dem trewen Schöpffer/ in gutten Wercken.*

Er redet hiemit an Patientes, die da leiden/ versteht aber dadurch nicht nur Speciatim, vnd in sonderheit die jehnigen/ so für andern viele vnd schwere Verfolgungen vmb GOTTes Lehre vnd Ehre willen/ vnd auch wol gar den Todt selbst leiden/ [...]

<https://www.deutschestextarchiv.de/359997/6>

Beispiel I: Charakteristische Sprachhandlungen

Imperativ in direktiven Sprachhandlungen

Du aber O HERre JEsu/ **gieb** vns hiezv
deine Gnade vnd guten Geist von oben/ auff
daß alles dir zu Ehren/ allen Betrübten zu kräft-
tigem Troste/ vnd ingesambt zu vnser aller See-
len Heyl vnnnd Seeligkeit gelange/ Amen.

<https://www.deutschestextarchiv.de/359997/11>



Imperativ wird hier im Rahmen
eines Gebets gebraucht, ist also stets
eine Bitte, *kein* Befehl, Anweisung o.
ä.

Denn **bedencket** die Vrsache/
warumb vnser lieber Herr den bitteren
Tod gelitten/ nemlich/ vmb vnser
Sünde willen. **Bedencket** auch die Liebe
Gottes/ daß er vns seinen Sohn ge-
schencket/ Da sehen wir Gottes Ge-
rechtigkeit vnd Barmhertzigkeit.

https://www.deutschestextarchiv.de/arndt_christentum01_1610/105



Imperativ ist hier Anweisung an die
Textrezipient*innen

Beispiel II: Nennung des Textproduzenten

Darumb/ sage **ich**/ muß des HERREN wort:
Jch werde nicht sterben/ in gewisser maß verstanden
werden/ nemlich/ nicht also sterben/ daß **Jch** im Tode
behalten werde.

<https://www.deutschestextarchiv.de/537789/10>

Jst so viel gesagt: Du/ mein Vater/
wirst es nicht zulassen/ daß **ich** im Grabe durch den Todt
vermodere/ sondern wirst mich im Grabe erhalten/ und
am dritten Tage Herrlich auferwecken.

<https://www.deutschestextarchiv.de/537789/11>

*Leitzitat der Auslegung: “Jch werde nicht sterben,
sondern leben und des HERREN Werck
verkündigen.” (Psalm 118, V. 17)*

Textproduzent

- wird selten einzeln genannt,
dann z. B. zur Emphase (wie
in diesem Fall)

Redewiedergabe (Bibelzitat)

- oftmals durch einleitende
Formel und “:”
gekennzeichnet

Fingierte Rede

- oftmals durch einleitende
Formel gekennzeichnet
(“Als wolle er sagen”)

Mixed Methods

Synthese beider Ansätze

Vergleich beider Methoden

Automatisierte Analyse

- + Große Textmengen können vollständig analysiert werden
- + Vergleich mehrerer Korpora möglich
- + Merkmale eines Phänomens können induktiv aus dem Datenmaterial ermittelt werden
- + die genaue Offenlegung der Untersuchungsmethode schafft Transparenz und Reliabilität und ermöglicht Reproduzierbarkeit
- Unterschiedliche Phänomene am gleichen Merkmal können oft nicht gut unterschieden werden
- durch den Fokus auf signifikant häufige Merkmale können weniger frequente, aber bedeutsame Phänomene aus dem Blick geraten

Manuelle Annotation

- Zeitaufwendig, daher kann nur ein Bruchteil der Korpora ausschnittsweise annotiert werden
- die Wahl der Annotationsebenen im Rahmen der Operationalisierung kann zur Einengung der Ergebnisse führen
- + Inhaltlicher Kontext kann bei der Merkmalsanalyse einbezogen werden
- + Annotationsmanuale führen zu möglichst validen, objektiven und reliablen Ergebnissen
- + die Annotationen können anschließend statistisch ausgewertet werden
- + aufbauend auf den Annotationen können automatische Klassifizierer trainiert werden

Verbindung manueller und automatisierter Methoden

Wie kann die manuelle Annotation textlinguistische Merkmale der Erbauungsliteratur erfassen, die mit der automatisierten Analyse nicht greifbar waren?

- **Wiederkehrende sprachliche Muster (z. B. potentielle Formeln)**
 - können ermittelt werden und enthalten ihrerseits Signalwörter, die wiederum in automatisierte Analyse einfließen können
- **Textstruktur**
 - bestimmte Textelemente folgen regelhaft aufeinander → diese Abfolgen können automatisiert als bestimmte funktionale Strukturen (z. B. Sprachhandlungen) identifiziert werden
- **Inhaltlicher Kontext**
 - Rückbezug auf bereits etabliertes Wissen → Ergänzung automatisierter Analyse, auch Rückführung in automatisierte Analyse möglich?

Ausblick

ToDoS Analyse

- Automatisierte Analyse auch für Zeitungen
- Auswertung der manuellen Analyse
- Zusammenführen der Ergebnisse der automatisierten und manuellen Analyse
- Vollständige Untersuchung des Textmusterwandels

ToDoS Datenaufbereitung

- DTABf und Zeitungsstrukturen
- Zusammenführen der manuell annotierten mit den DTABf-Dokumenten