

t.evo - Die Evolution von komplexen Textmustern

Ein DFG-gefördertes Kooperationsprojekt der Universität Paderborn
und der Berlin-Brandenburgischen Akademie der Wissenschaften

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

Überblick Digitalisierungsworkflow

Korpora:

Zeitungen – (Arbeitsstelle Paderborn)

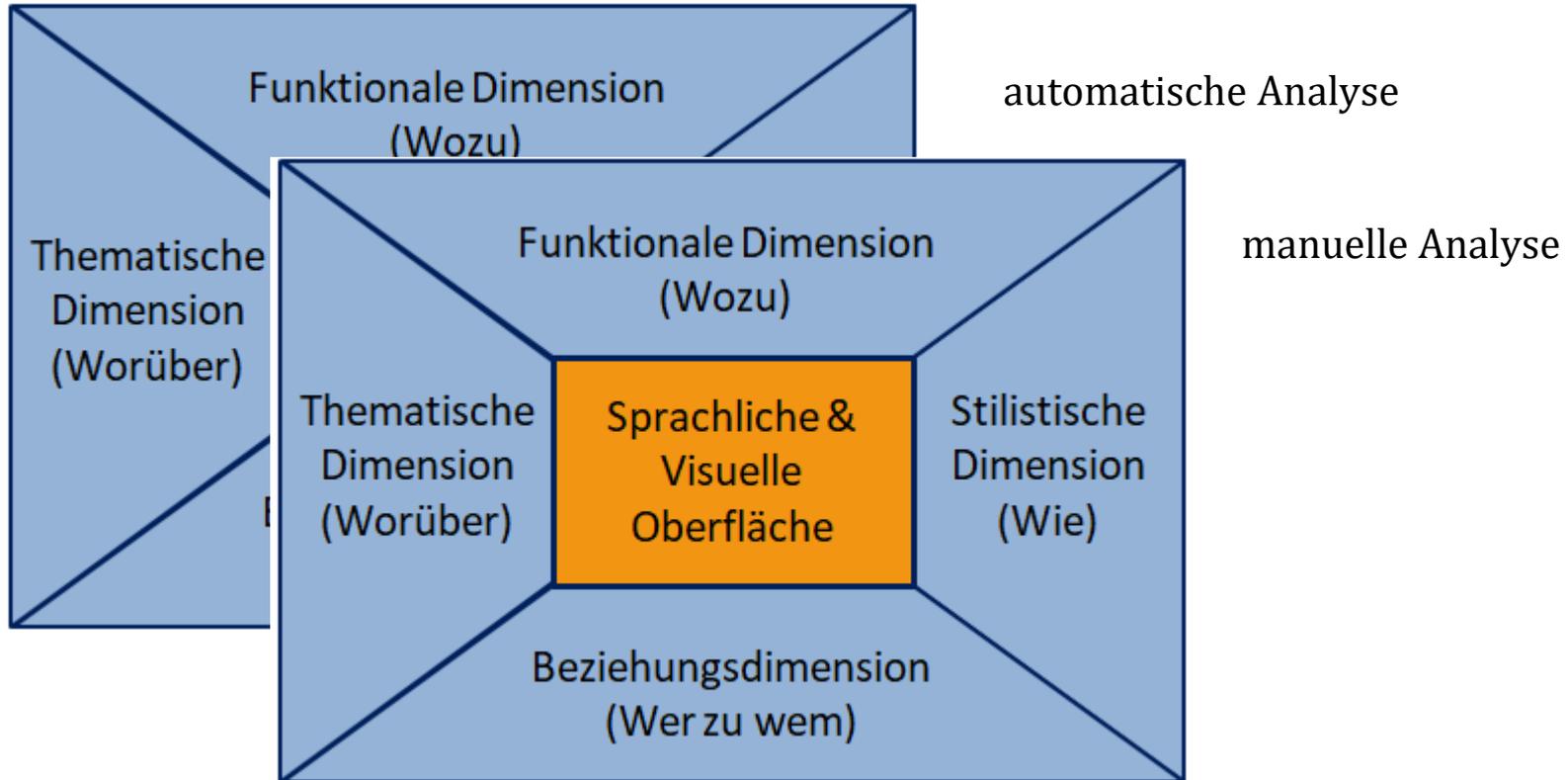
- 50 Ausgaben der Allgemeinen Zeitung zur manuellen Annotation (1830–1929)
- 125 weitere Ausgaben der Allg. Zeitung für quantitative Analysen

Erbauungsliteratur – (Arbeitsstelle Berlin)

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

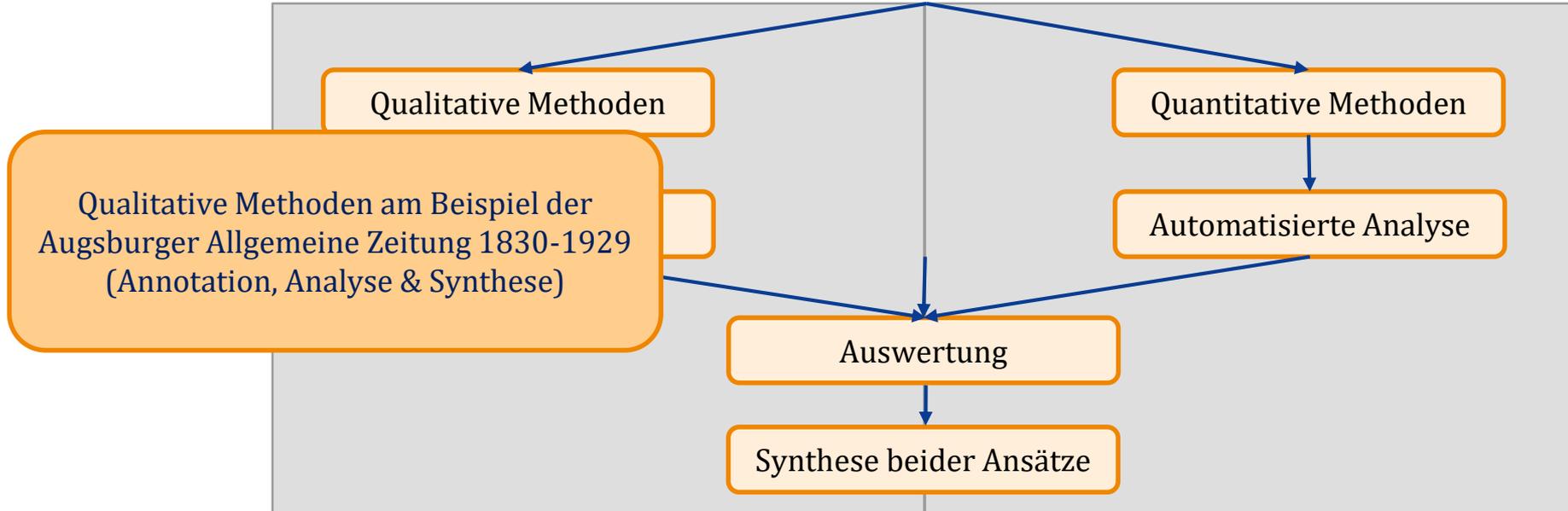
Unterscheidung Annotation und quantitative Analyse für beide Korpora

Analysemodell in t.evo



Vom Faksimile zum Textmuster

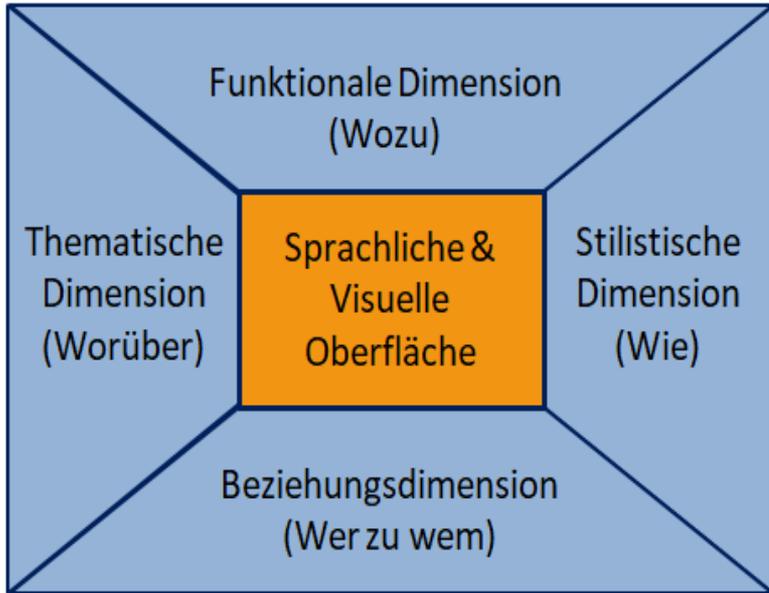
1. Erstellung der Korpora – Ausgewogenheit und Repräsentativität
2. Digitalisierung – einheitlich codierte Datengrundlage (DTABf, TEI-Codierung)
linguistisch analysierte Volltexte (Lemma, POS, Satzgrenzen)



Alles, wie es immer war?

Algier, 21 Dec.* Die Gerüchte, welche sich in unserer Stadt von einem bei Maison carrée vorgefallenen Gefecht, in dem die Araber 3 bis 4000 Mann verloren haben sollen, verbreitet hatten, waren falsch. Der große Sieg reducirt sich auf ein ganz unbedeutendes Resultat. Wir hatten zwar gleich vermuthet, daß jene Gerüchte übertrieben seyen, dachten aber doch, man habe den Arabern bei Maison carrée eine tüchtige Lection gegeben. Es blieben aber nur etwa hundert Araber auf dem Kampfplatz, wie uns Officiere, die aus der Metidscha kommen, versichern. Der Marschall Valée scheint trotz der bedeutenden Verstärkungen, die er erhalten, noch nicht Willens, vorzurücken, und der Krieg wird wohl bis zum künftigen Frühjahr nur innerhalb der Metidscha geführt werden. Die Verstärkungen, welche die letzten Schiffe aus Frankreich brachten, betragen über 3000 Mann. Auf dem Abhang des Sahel steht jetzt eine furchtbare Vertheidigungslinie und alle Lager haben Verstärkungen erhalten, um Ausfälle zu machen. Uebrigens ist unsere Lage fortwährend dieselbe. Die Lebensmittel sind theuer und die Märkte leer.

Dimensionen und Annotationsebenen



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

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

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

→ Ebene 1 - Quellspezifische Tags: Textsorten der Pressekommunikation & pressspezifische 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)

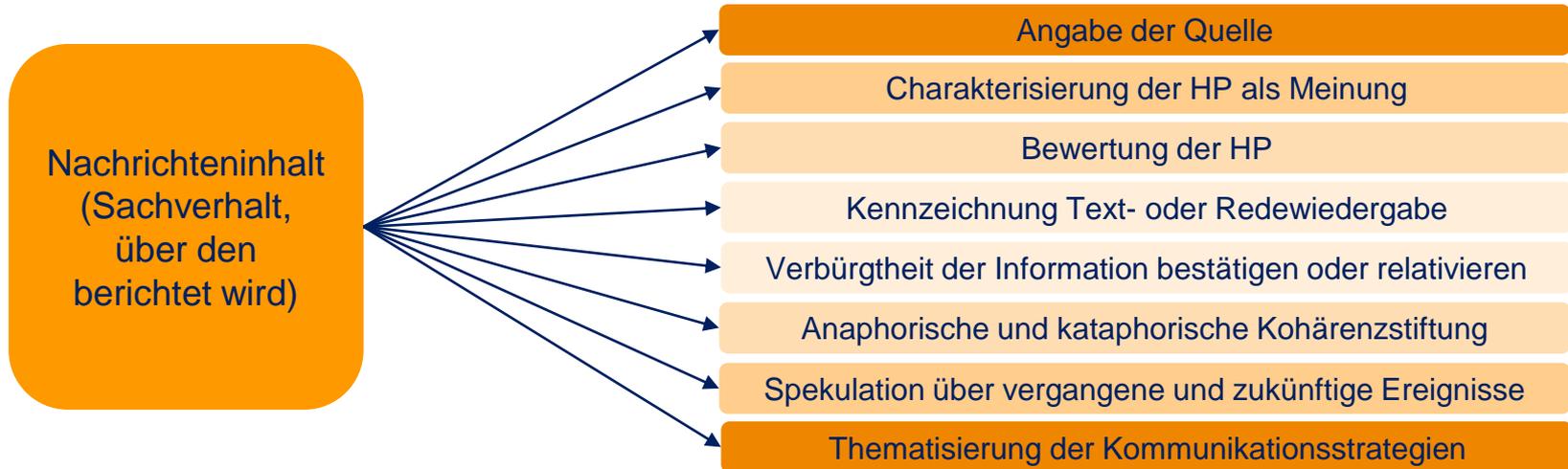
Ebene 5 - Thematische Dimension: "journalistische Ressorts" (12 Tags)

Manuelle Annotation - 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

Anwendungsbeispiel: 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.



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.

Anwendungsbeispiel: Potentielle Formeln

Ziele der Annotation:

1. Prädikationstypen und Presstextsorten: Zusammenhang zwischen Presstextsorten und der Verwendung bestimmter (formelhafter) sprachlicher Einheiten (Nebenprädikationen) (Bsp. 1)
2. Musterbildung und Validierung am Gesamtkorpus: Ergebnisse der manuellen Annotation für Erhebung dieser (formelhaften) sprachlichen Einheiten durch automatisierte Verfahren nutzen (Bsp.2)

Bsp. 1: Prädikationstypen und Presstextsorten

"Die Differenz zwischen Nachrichtenbericht/Meldung und Rasonnement bei den journalistischen Berichten liegt also anfangs keineswegs in der Differenz zwischen Sachlichkeit der Berichterstattung und Meinung im Kommentar, sondern in der isoliert ereignisbezogenen und auf Glaubwürdigkeit bewerteten Mitteilung («Tagesneuigkeit») und der allgemeineren (meinungszentrierten) Darstellung von Sach- und Rechtsverhältnissen." (Ramge 2008: 34)



- Angabe der Quelle einer Information
- Verbürgtheit der Quelle bestätigen oder relativieren
- Bewertung der Hauptprädikation als Meinung

Formelhafte Sprache und manuelle Annotation

„Wenigstens verfichern zwei Briefe aus Czernowitz (Bukowina) vom 24 und von der moldauisch-itebenbürgischen Gränze vom 20 März“

der „Presse“ zufolge

(Privattelegramm.)

Ein Schreiben aus Astrachan

- sagt die Weserzeitung -

Angabe der Quelle einer Information

Nach der Colonial Gazette

Es heißt

Die stenographischen Berichte, die jetzt vorliegen,...

Man fagt auch

ein Parifer Blatt

(Münchn. pol. Z.)

‘... ,wie die „Pol. Corr.“ meldet, ...

Formelhafte Sprache und manuelle Annotation

„Die Meldung ist auch in diesem wichtigsten Punkte falsch gewesen.“

Die stenographischen Berichte, die jetzt vorliegen, lassen keinen Zweifel

Der „Observer“ erhält aus „vertrauenswürdiger Quelle“ die Mittheilung

wir sind jedoch nicht geneigt, diesem Gerüchte Glauben beizumessen

Verbürgtheit der Quelle bestätigen oder relativieren

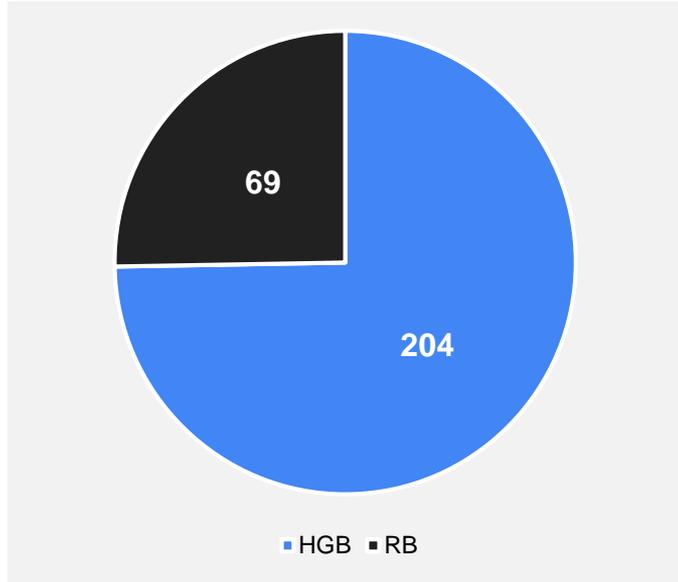
In den diplomatischen Cirkeln wird indeffen verifiziert

Durchaus willkürlich und unberechtigt ist übrigens die Unterstellung Brüsseler Blätter

Die Meldung ist auch in diesem wichtigsten Punkte falsch gewesen.

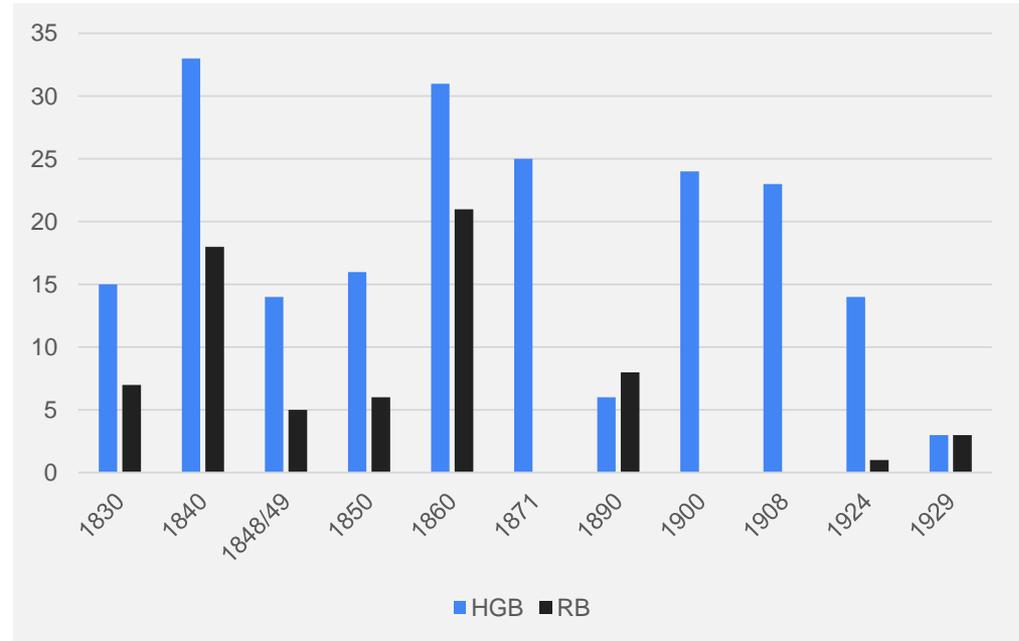
Wir können die Zuverlässigkeit dieses Briefs verbürgen.

Formelhafte Sprache und manuelle Annotation



Verteilung der Angaben zur Quelle und deren Verbürgtheit nach Textsorten

Verteilung der Angaben zur Quelle und deren Verbürgtheit nach Textsorten und Jahrgängen



Formelhafte Sprache und manuelle Annotation

„Dieß sind Thatsachen, welche jedes Raisonement überflüssig machen.“

Wir sind der Meinung, daß

unterliegt der Natur der Sache nach wohl keinem Zweifel.

Doch ist nach unfreier Ueberzeugung alle Hoffnung vorhanden

Bewertung der Hauptprädikation (Textproduzent)

ich habe Grund zu vermuthen

Es ist zweifellos

Und noch stärker in Zweifel zu ziehen ist es

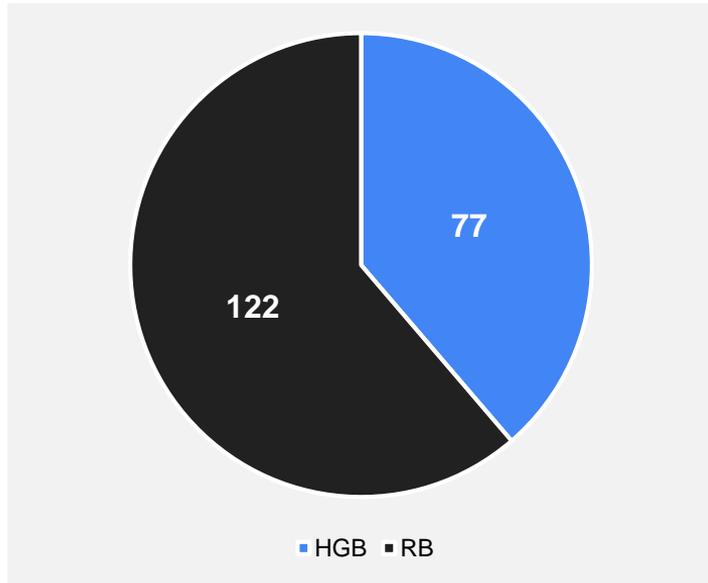
All das heißt natürlich nicht, daß wir nun glauben dürften, die Dinge nähmen von jetzt an einen günstigen Verlauf.

Wir glauben

Es ist unrichtig

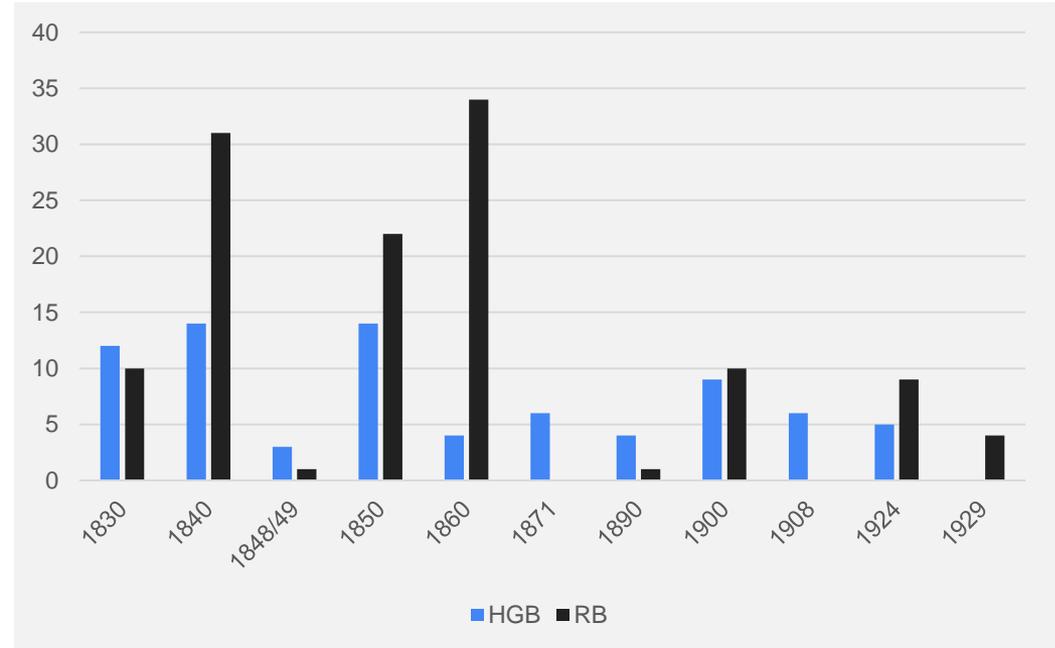
In unsrenen Augen

Formelhafte Sprache und manuelle Annotation



Verteilung von Bewertungen der Hauptprädikation nach Textsorten

Verteilung von Bewertungen der Hauptprädikation nach Textsorten und Jahrgängen



Formelhafte Sprache und manuelle Annotation

- Fazit:

- Erfassung aller Zusatzprädikationen unter dem Tag „Potentielle Formel“ sinnvoll, da Prädikationstypen häufig in Kombination verwendet werden, z.B.:

Quellenangabe + Glaubwürdigkeit

nur trüber Quelle entstehenden Gerüchten

Quellenangabe + Redeeinleitung

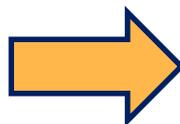
„Narodni Lifty“ sagen zwar,...

- Zusammenhang zwischen Presstextsorte und der Verwendung von „Potentiellen Formeln“ des Typs „Angabe der Quelle“, „Glaubwürdigkeit einer Quelle bestätigen oder relativieren“ sowie „Bewertung der Hauptprädikation als Meinung“ lässt sich auf Basis der verwendeten Texte(21 Ausgaben) nachzeichnen

Bsp. 2: Musterbildung und Validierung am Gesamtkorpus

„Work in progress“

1 von Grund ...	(Hannover'sche Ztg.)	Oesterreich.
wsk, 1. April...	(Meldung der R[...]hen-Agentur.)	Nach erfolgt
id Schöning...	Allein über da[...] zu erheben,	denn fichere
les gesamt...	Angefangen sin[...]eiten schon:	von Ans bis l
. K. C.	Aus Bombay wir[...] berichtet:	In Peichauer
T. N.)	Aus Rom, 2 Mär[...] geschrieben:	Gestern übe
heil- scheine.)	Bei der heute [...]ien gezogen:	16, 1002 unc
ft zu machen.	Das „ Siècle “[...]t Entrüstung:	„Kein Wort d
Jeutlichen St...	Dem Vernehmen nach	haben die M
erlin, 11 März.	Dem Vernehmen nach	ist die Demc
urch die Eng...	Der Fall wird [...]olgt erzählt:	Am 13 Aug. l
le in Bresla...	Der Kultusminifter erklärt,	diefe Idee fei
he, 2. April.Tel.	Der badische H[...]liche befragt,	daß der Han
stag ausgeg...	Derfelbe behan[...]en, nämlich:	1. Anzeigepef
zu gewäh- ren.	Die Blätter bemerken,	dies fei nach



Muster:	es (sein VAFIN) ADJD	Abs. Freq.:	1	Rel. Freq.
Datei:	AAZ_08_09_01_1924.xml			
Muster:	es (sein VAFIN) ADJD	Abs. Freq.:	4	Rel. Freq.
Datei:	AAZ_135_21_03_1908.xml			
Muster:	es (sein VAFIN) ADJD	Abs. Freq.:	3	Rel. Freq.
Datei:	AAZ_43_24_10_1914.xml			
Muster:	es (sein VAFIN) ADJD	Abs. Freq.:	4	Rel. Freq.
Datei:	AAZ_75_15_03_1848.xml			
Muster:	es (sein VAFIN) ADJD	Abs. Freq.:	2	Rel. Freq.
Datei:	AAZ_11_14_01_1929.xml			
Muster:	es (sein VAFIN) ADJD	Abs. Freq.:	6	Rel. Freq.
Datei:	AAZ_04_04_01_1872.xml			
Muster:	es (sein VAFIN) ADJD	Abs. Freq.:	9	Rel. Freq.
Datei:	AAZ_05_05_01_1830.xml			
Muster:	es (sein VAFIN) ADJD	Abs. Freq.:	4	Rel. Freq.

Vollständige Sätze des Musters mit Metadaten:

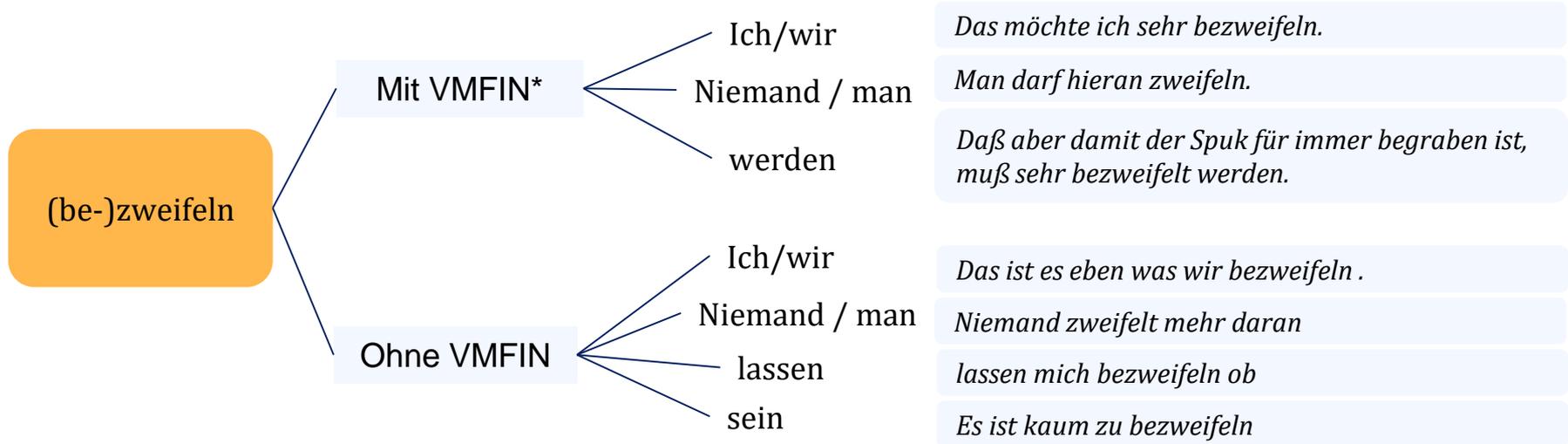
Datei: AAZ_137_23_03_1908.xml
 Sentence: Es ist wahrscheinlich , daß auch die Kattune der Waschkleider mit ähnli mehr , als das Bestreben vorhanden ist , der Kattundruckerei , die in der zweiten Jony hoffähig war , zu einem neuen Aufschwung zu verhelfen .
 Sentence: Es ist selbstverständlich , daß man eine Schar von Künstlern nicht so di:
 Datei: AAZ_38_19_09_1914.xml
 Sentence: Es ist richtig , wir haben Belgiens Neutralität verletzt , weil uns die bitt

KWIC-Anzeige Potentielle Formeln in Catma 6

Ergebnis: Suchanfrage es (sein | VAFIN) ADJD

Musterbildung und Validierung am Gesamtkorpus

„An ihrer Zustimmung ist jedoch nicht zu zweifeln.“

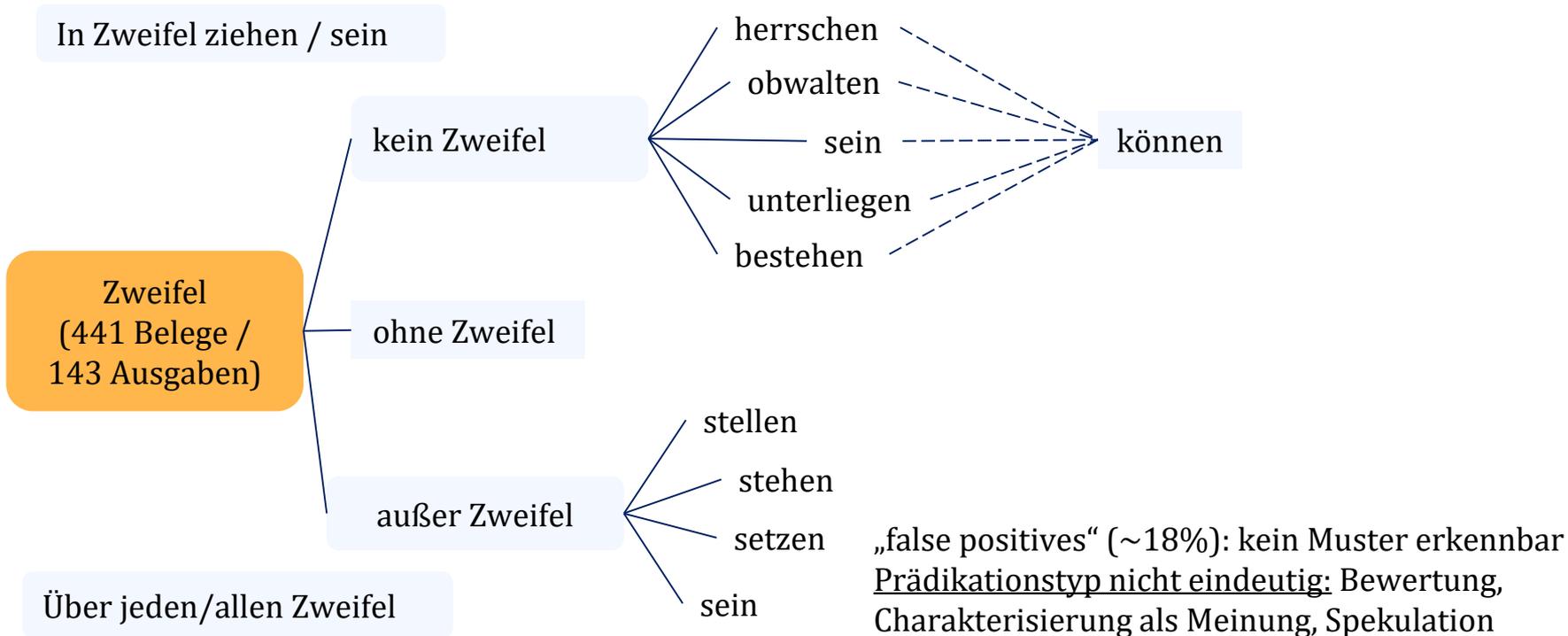


„false positives“ (~20%): Pron (er/sie) + 3. Pers. Sg. (!), Präteritum
Prädikationstyp nicht eindeutig: Bewertung der Hauptprädikation,
Charakterisierung als Meinung, Spekulation

*VMFIN: „können“, „dürfen“, „müssen“, „möchten“

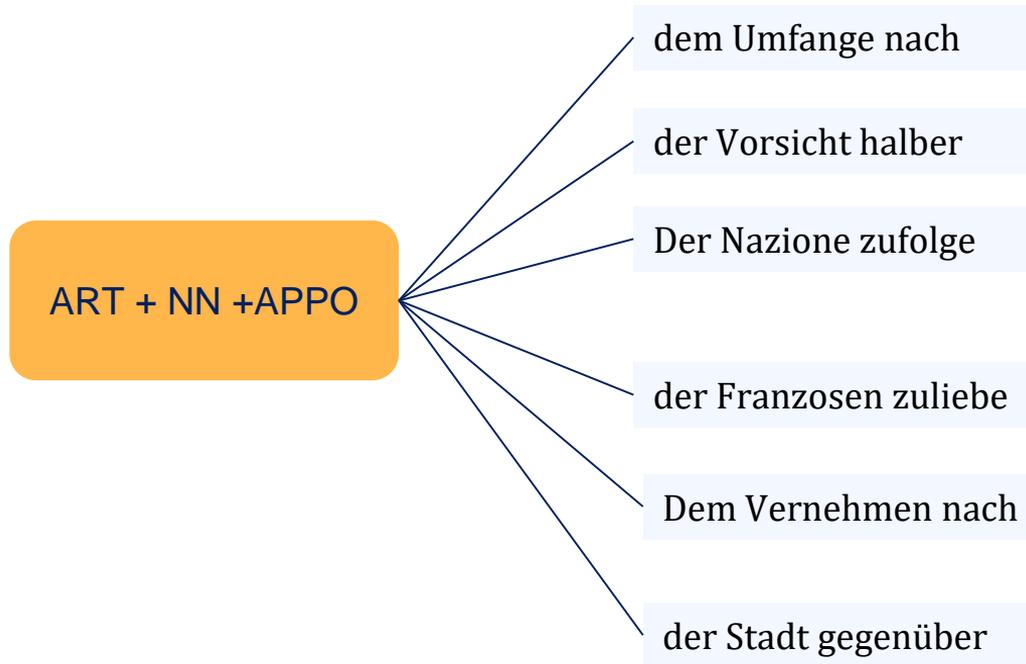
Musterbildung und Validierung am Gesamtkorpus

„Die gedachte Denkschrift wird ohne Zweifel darüber allen Aufschluß ertheilen.“

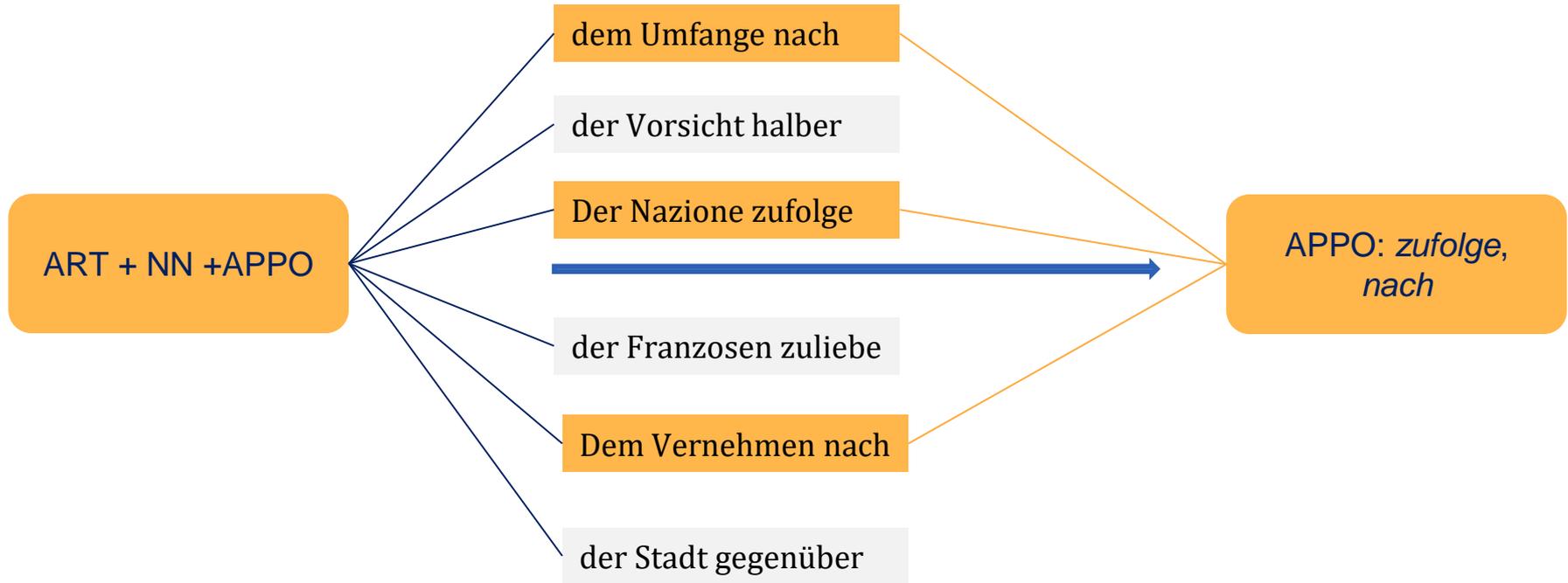


Musterbildung und Validierung am Gesamtkorpus

„Einem Gerücht zufolge werde die Frage in die Phase der Unterhandlungen eintreten.“



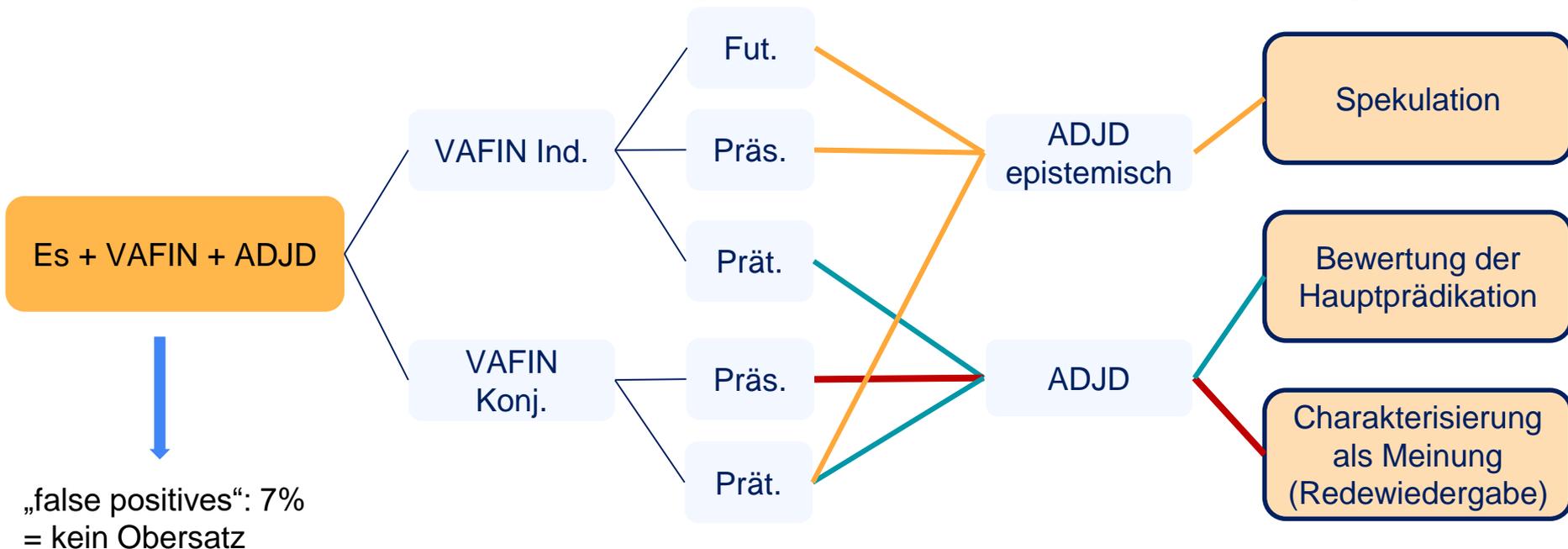
Musterbildung und Validierung am Gesamtkorpus



„Einer bei wohlunterrichteten Leuten ziemlich verbreiteten Meinung zufolge“

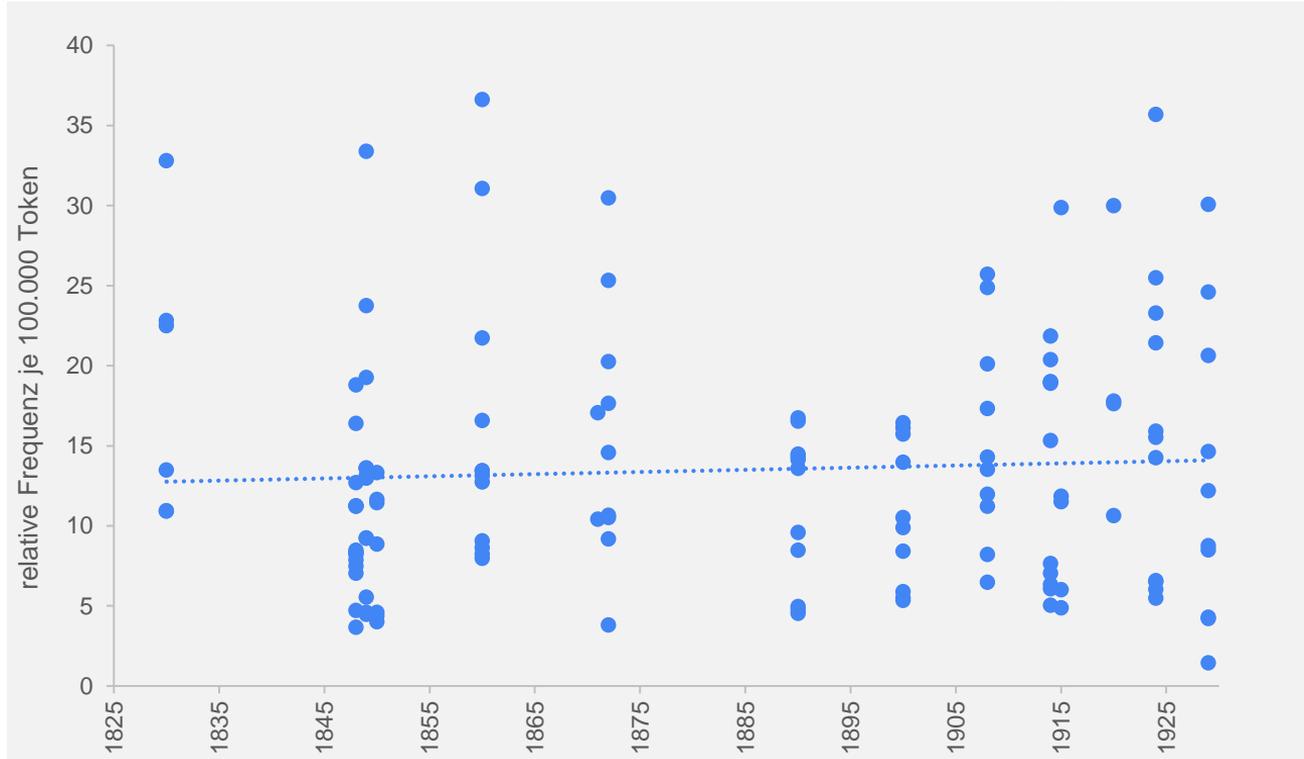
Musterbildung und Validierung am Gesamtkorpus

„Aber es ist besser, einen Fehler einzugestehen, als ihn ein zweites Mal zu begehen.“



Musterbildung und Validierung am Gesamtkorpus

Es VAFIN ADJD (feste Abstände) im Gesamtkorpus (relative Frequenz je 100.00 token)



Fazit

- Teil 1:
 - Ein Zusammenhang zwischen Presstextsorte und verwendeten Nebenprädikationen (formelhafte Wendungen) lässt sich anhand der manuellen Annotation nachweisen
 - Die Erfassung aller Nebenprädikationen, die formelhafte Wendungen ausbilden könnten, erweist sich aus zwei Gründen als sinnvoll:
 - Überschneidung verschiedener Prädikationen
 - Musterbildung nicht von Prädikationstyp abhängig
- Teil 2:
 - Es lassen sich musterhafte, abfragbare Strukturen und Verwendungsweisen erkennen in
 - Lexik
 - Grammatik
 - „Nachjustieren“ der bisher abgefragten Muster notwendig
 - Muster, die Lexik und Grammatik miteinander verbinden, zeigen höchste Trefferquoten
 - Tendenzen und Entwicklungen auf derzeitiger Datenbasis noch nicht möglich

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?