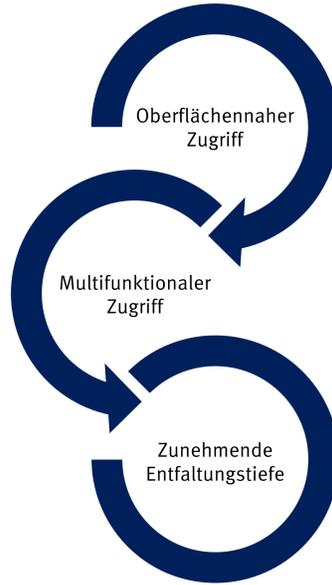


MERKMALE REGISTRIEREN ODER TEXTUELLE PHÄNOMENE IDENTIFIZIEREN?

ZUR VEREINBARKEIT VON AUTOMATISCHER UND MANUELLER TEXTSORTENANALYSE

Phänomenorientierter Ansatz

Grundsätzlich werden Texte in der linguistischen Textanalyse als Vertreter einer Textsorte betrachtet; der Textsortenbegriff ist stark spezifiziert. Textsorten werden mit sich wiederholenden kommunikativen Aufgaben identifiziert, für die konventionelle und muster-hafte Lösungen in unterschiedlichen Kommunikations-domänen gefunden wurden. Sie gelten als historisch ver-änderbare und kontextuell eingebettete Instrumente sozialen Handelns. Die linguistische Textanalyse folgt dem Gedanken der Mehrdimensionalität von Texten und registriert Textmusterrealisierungen in der funktionalen, thematischen, situativen und sprachlich-stilistischen Dimension. Von diesem Hintergrund aus wird die jeweilige sprachliche Einheit in Hinsicht auf ermittelte Dimensionen und auf die ihnen zugrundeliegenden Aufgaben interpretierbar, d.h. sprachlichen Formen können entsprechende Funktionen zugeordnet werden. Dies gilt auch für die holistische Kategorie des Textsortenstils.



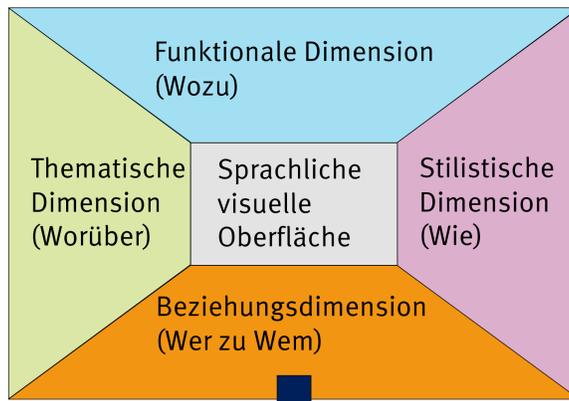
Beziehungsdimension

Qualitatives Vorgehen

Für die manuelle Annotation in Catma wurde basierend auf einschlägigen Forschungsergebnissen ein Tagset erstellt, das als agiles Tagset konzipiert ist und so den besonderen Anforderungen und „Textrealitäten“ angepasst werden kann. Diese Anpassung erfolgt in einem mehrstufigen Verfahren, wie es im Qualitätsmanagement (PDCA-Methodik) üblich ist:

Aufbau eines Tagsets	(Plan);
Anwendung auf Pilotkorpus	(Do);
Auswertung und Problemerkennung	(Check);
Anpassung des Tagsets und Anwendung auf Korpus	(Act).

Bezogen auf das Fallbeispiel bedeutet dies eine Differenzierung nach Textproduzent, Textrezipient, Fremdgruppe und Informant, wodurch einerseits die Gesamtheit der verwendeten Formulierungen erfasst wird, andererseits deutlich wird, dass Personaldeiktika wie *wir* auf verschiedene Personengruppen angewendet werden.



Exemplarische Analyse personaldeiktischer Ausdrücke:
ich - du - wir - ihr

Merkmalsorientierter Ansatz

Die Kategorie der Textsorte stellt in der Regel eine Hilfskategorie dar, mit der größere Datenmengen (z.B. Referenzkorpora) geordnet werden. Texte werden zudem für das Training von Methoden (z.B. der Sentiment Analysis oder des Topic Modeling) verwendet. Der Text(sorten)begriff bleibt dabei unterspezifiziert, indem Text mit Dokumenten, Sätzen oder Mengen sinntragender Struktureinheiten gleichgesetzt oder nach dem Alltagsverständnis differenziert wird. Das Text- und insbesondere das Stil- und Registerverständnis sind vor diesem Hintergrund wesentlich an Merkmalen orientiert. Zu Merkmalen zählen u.a. Frequenzen von Inhalts- oder Funktionswörtern, Variantenreichtum des Wortschatzes, Satztlängen oder n-Gramme. Ihre Auswahl scheint durch ihre Operationalisierbarkeit selbst gerechtfertigt. Auch multidimensionale Analysen, die eine Vielzahl von Merkmalen zu Merkmalsbündeln zusammenfassen, berücksichtigen kaum deren interpretierbare Kategorien – etwa die Funktionalität der sprachlichen Formen. Primärer Bezugspunkt bleibt die Textoberfläche.

Beziehungsdimension

Quantitatives Vorgehen

Die Basis für die Analyse bilden zwei Korpora aus Texten der Erbauungsliteratur sowie aus vollständigen Zeitungsausgaben, die jeweils um ein Vergleichskorpus ergänzt werden. Die Korpora sind dabei dem DTA-Gesamtkorpus entnommen.

Im Rahmen der Analyse personaldeiktischer Ausdrücke werden verschiedene Personal- und Possessivpronomina als Merkmale ausgehend von XPath-Suchausdrücken erfasst und gezählt:

```
//w[@lemma='wir'][@pos='PPER']
```

Zur Visualisierung der Ergebnisse werden Streu- und Kastendiagramme genutzt. Diese geben Auskunft darüber, wie sich die Häufigkeit eines Merkmals über die Zeit im Durchschnitt verändert und wie hoch die Abweichung für ein Merkmal im Korpus ist. Die Darstellungen eignen sich in besonderer Weise zur Identifikation von Trends im Rahmen des historischen Wandels bezogen auf Textsorten, die kulturindizierend wirken können. Dabei können auch Ausreißer interessant sein. Diese bedürfen der qualitativen Deutung.

Beispiele – erste Analyseergebnisse: qualitativ und quantitativ

(1) AZ 1840 01 30; Schwedische Zustände. II. Beamtenbildung. Stockholm, Januar.: „Wie glücklich, würden **wir Deutschen** hiebei ausrufen können, [...]“

(2) AZ 1840 06 29; Dresden, 20. Juni: „**Meine Herren Stände!** Aufs neue stehen **wir** am Schluß eines wichtigen Abschnittes in unserm Staatsleben [...]“

(3) AZ 1840 03 18; Ein Besuch beim Vladika von Montenegro: Man schätzt die Höhe des Monte Sella (Lovczin) auf 6000 Fuß über der Meeresfläche, doch scheint mir diese Angabe übertrieben; der Bergkamm, **den wir überstiegen**, dürfte zwischen 4000 und 5000 Fuß hoch seyn.

(4) AZ 1840 02 08; London, 30. Jan.: „Sie sind alle, **wie wir hören**, in guter Gesundheit, und mit der Behandlung, die sie von dem Capitän erhalten haben, wohl zufrieden.“

(5) AZ 1840 06 23; Preußen. Vom Niederrhein. 16. Juni: „Es ist **Unser** fester Wille, unter Gottes Beistand, Recht und Gerechtigkeit zu handhaben, und das Wohl **Unserer** Unterthanen aus allen Kräften zu fördern. **Wir** erwarten von den Mitgliedern **Unserer** Regierung, daß sie dieser **Unserer** Absicht ihrerseits durch Berufstreue und Gewissenhaftigkeit in aller und jeder Weise entsprechen werden. [...]“

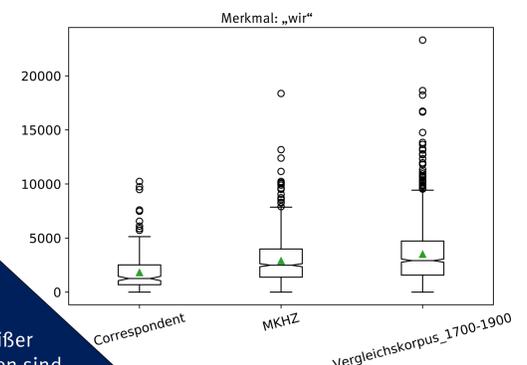
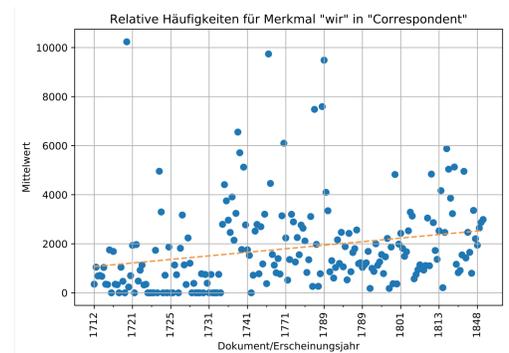
Das Bedeutungsspektrum von *wir* reicht nach dem DWB von der Grundbedeutung der „1. person. in der mehrzahl, für den engeren (2+3) oder weiteren kreis (1), zu dem sich der sprechende oder schreibende rechnet“, bis hin zur Bezeichnung eines Verfasserkollektivs (4) oder einzelner Personen (5).

Es bleibt also ohne Blick in den Kontext zunächst völlig offen, wer *wir* sind oder ist. Die in (1)-(5) bezeichneten Personenkreise (PK) werden immer enger gefasst, d.h. vom alle inkludierenden Gebrauch von *wir* in (1) bis hin zum alle exkludierenden Gebrauch in (5):

- (1) *wir + Deutschen*: PK, der über Textproduzent und -rezipient hinausgeht (alle Deutschen)
- (2) *Meine Herren Stände + wir*: PK der Textproduzenten und direkten Rezipienten umfasst, jedoch nicht den Leser
- (3) *den wir überstiegen*: Reisegruppe, zu der der Verfasser gehört, die aber keinen Rezipienten einschließt
- (4) *wie wir hören*: Verfasserkollektiv, das das Gehörte der Leserschaft vermittelt
- (5) *Wir / Unsere Unterthanen*: Rede des preußischen Königs, der von sich selbst als *wir* spricht

Textsortenspezifik:

Das Streudiagramm zeigt einen Aufwärtstrend bei dem Pronomen *wir* im historischen Verlauf. Die Punkte innerhalb des Kastendiagramms weisen auf Ausreißer im Korpus hin, die im Rahmen einer weiterführenden qualitativen Analyse zu überprüfen sind. (1)-(4) repräsentieren für die Zeitungskommunikation typische Fälle. Dabei kann einerseits eine Gemeinschaft mit dem Leser oder darüber hinaus gebildet werden (1)+(2), besonders häufig jedoch findet man Formen von (3) und (4), die der Verifikation von Nachrichten oder dem Erfahrungsbericht entstammen. Verwendungsvariante (5) hingegen folgt alten Formulierungstraditionen (pluralis majestatis) und ist in der Pressekommunikation bestenfalls typisch für die wörtliche Wiedergabe von Reden.



Literatur: Adamzik, K. (2016²): Textlinguistik. Grundlagen, Kontroversen, Perspektiven. Berlin, Boston; Bubenhofer, N./Dreesen, Ph. (2018): Linguistik als antifragile Disziplin? Optionen in der digitalen Transformation. In: Digital Classics Online 4 (1): S. 63–75; Bubenhofer, N./Scharloth, J. (2015): Maschinelle Textanalyse im Zeichen von Big Data und Data-driven Turn – Überblick und Desiderate. In: Zeitschrift für Germanistische Linguistik 43 (1): S. 1–26; Sandig, B. (2006²): Textstilistik des Deutschen. Berlin; Haaf, S. (2019): Mehrdimensionale Beschreibung erbaulicher Textsorten des 17. Jahrhunderts. Ein korpusbasierter Ansatz. In: Szczepaniak, R./ Hartmann, S./ Drücker, L.: Historische Korpuslinguistik. Berlin, Boston: S. 324–344.

