

Prof. Dr. Andreas Blätte (Universität Duisburg-Essen) | Prof. Dr. Gary Schaal (Helmut-Schmidt-Universität)

# Plenardebatten als öffentliche Sprachressource der Demokratie

## Kurationsprojekt F-AG 8 (Inhaltsanalytische Methoden in den Sozialwissenschaften)

### Das Kurationsprojekt in Kürze

- Bestehende Ressource: PolMine-Plenarprotokollkorpus (1996 – 2013, vgl. www.polmine.de)
- Im Kurationsprojekt: Annotation und Klassifikation von Plenardebatten
- Durch Möglichkeit der Bildung themenspezifischer Subkorpora Erweiterung der Einsatzmöglichkeiten des Korpus
- Gewährleistung des Datenzugangs durch entsprechende Infrastruktur / Hosting-Lösung

### Ressource

#### Warum ein Plenarprotokollkorpus?

- Plenarprotokolle repräsentieren die volle Bandbreite des politischen Geschehens
- Sie werden von Parlamenten kontinuierlich produziert - über die Parlamentsarchive sind sie für größere Zeitspannen verfügbar
- Lizenzrechtlich sind Plenarprotokolle gemeinfrei, der Nachnutzung eines Plenarprotokollkorpus stehen keine rechtlichen Hürden entgegen

#### Stand der Ressource: Annotation

Die Plenarprotokolle des Bundestags wurden im PolMine-Projekt in einem automatisierten Verfahren „XMLifiziert“. Das Plenarprotokollkorpus-XML enthält als Metainformationen bzw. Annotationen:

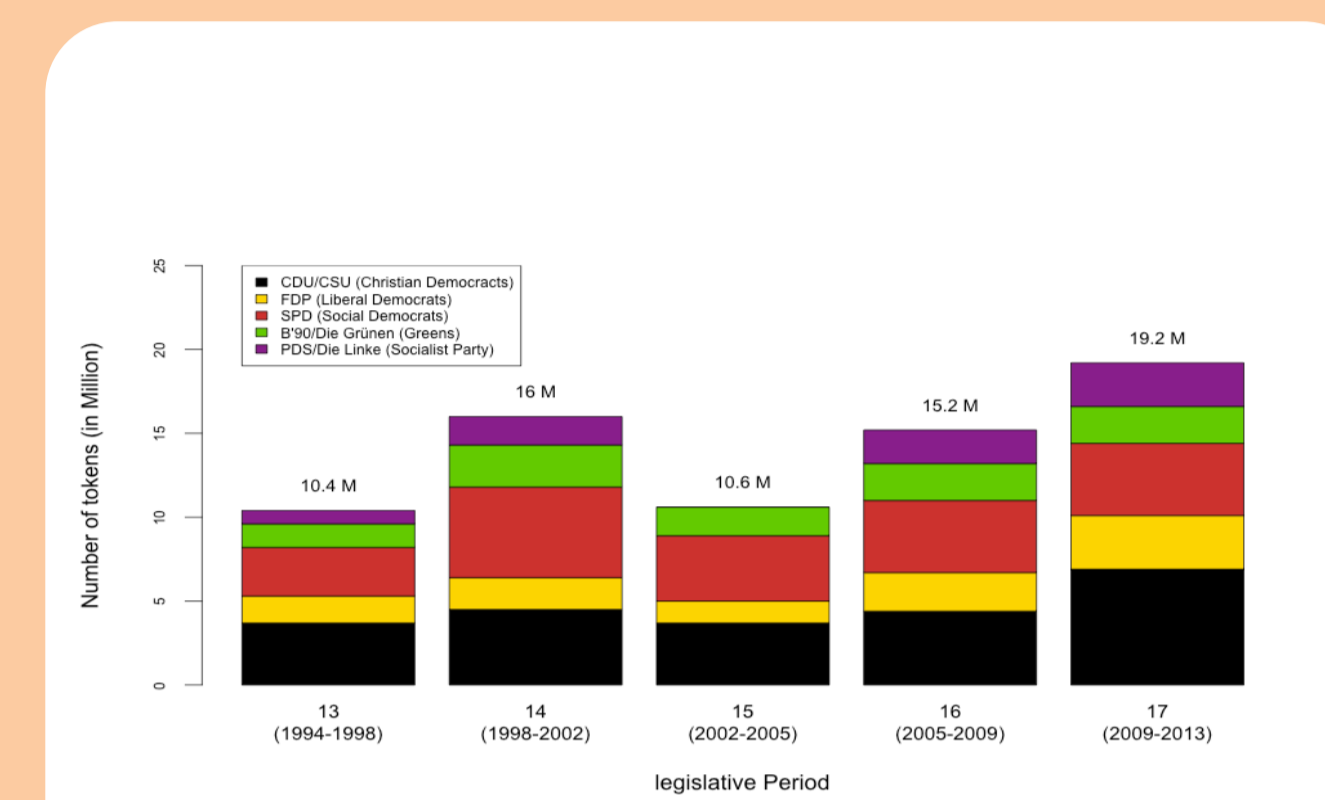
- Zeitstempel / Datum einer Debatte
- Annotation von Rednern
- Partei- bzw. Fraktionszugehörigkeit von Rednern
- Annotation von Zwischenrufen

Das Korpus ermöglicht so die Bildung von Teilkorpora nach variablen Kriterien.

#### Datenbestand: Überblick

| WP | Bestand        | Zeitraum              | #Protokolle | #Token     |
|----|----------------|-----------------------|-------------|------------|
| 13 | ab 86. Sitzung | 08.02.1996-16.10.1998 | 163         | 11.484.628 |
| 14 | vollständig    | 26.10.1998-13.09.2002 | 253         | 18.955.237 |
| 15 | vollständig    | 17.20.2002-28.09.2005 | 187         | 12.797.634 |
| 16 | vollständig    | 18.10.2005-08.09.2009 | 233         | 17.623.703 |
| 17 | vollständig    | 27.10.2009-03.09.2013 | 251         | 22.544.458 |
|    |                | 08.02.1996-03.09.2013 | 1087        | 83.405.660 |

#### Datenbestand (nach Parteien)



### Klassifikation

#### Kurationsprojekt: Arbeitspakete

- Annotation von Tagesordnungspunkten in XML-Plenarprotokollen
- Bestimmung eines Klassifikationsschemas
- Definition von Codierregeln und Coderschulung
- Erstellung von Trainingsdaten (manuelle Codierung von Plenardebatten)
- Klassifikation des vollständigen Datensatzes auf Basis der Trainingsdaten
- Bereitstellung der aufgewerteten Ressource (s.u.)

#### Annotation von Tagesordnungspunkten

Im PolMine-Plenarprotokollkorpus sind Tagesordnungspunkte bislang nicht annotiert. Die Annotation erfolgt im ersten Arbeitspaket.

**Vizepräsidentin Eitelgard Reimann:**  
Liebe Kollegen und Kollegen, ich schließe diese...  
...  
Interaktionell wird die Überweisung der Vorlagen...  
...  
Drucksache 18274

**Vizepräsidentin Uta Schmidt:**  
Vielen Dank... ich schließe die Aussprache...  
...  
Tagesordnungspunkt 25 auf...  
...  
Wahlrecht - Bildung und Gesundheit...  
...  
Nach einer interaktiven...  
...  
Ich öffne die Aussprache...  
...  
Ich bin bei der CDU/CSU

#### Klassifikationsschema

Ausgangspunkt für das Projekt ist das Schema des Comparative Agendas Project (CAP) als gut etablierte politikwissenschaftliche Differenzierung von Politikbereichen.

| #no | Beschreibung (englisch)                                |
|-----|--|
| 1   | Domestic Macroeconomic Issues                          |
| 2   | Civil Rights, Minority Issues, and Civil Liberties     |
| 3   | Health   |
| 4   | Agriculture  |
| 5   | Labor and Employment                                   |
| 6   | Education  |
| 7   | Environment  |
| 8   | Energy   |
| 9   | Immigration and Refugee Issues                         |
| 10  | Transportation   |
| 11  | Law, Crime, and Family Issues                          |
| 12  | Social Welfare   |
| 13  | Community Development and Housing Issues               |
| 14  | Banking, Finance, and Domestic Commerce                |
| 15  | Defense  |
| 16  | Space, Science, Technology, and Communications         |
| 17  | Foreign Trade  |
| 18  | International Affairs and Foreign Aid                  |
| 19  | Government Operations                                  |
| 20  | Public Lands, Water Management, and Territorial Issues |
| 21  | Cultural Policy Issues                                 |

#### Erweiterte Auswertungsmöglichkeiten

Das um die Klassifikation erweiterte Korpus ermöglicht eine Reihe themenspezifischer Auswertungsmöglichkeiten zu den klassifizierten Politikbereichen, etwa zu:

- Wandel politischer Sprache
- Themenkonjunkturen
- politischer Wettbewerb um Schlüsselbegriffe

Das Korpus ist über die Politikwissenschaft hinaus relevant für jede Beschäftigung mit parlamentarischen Diskursen.

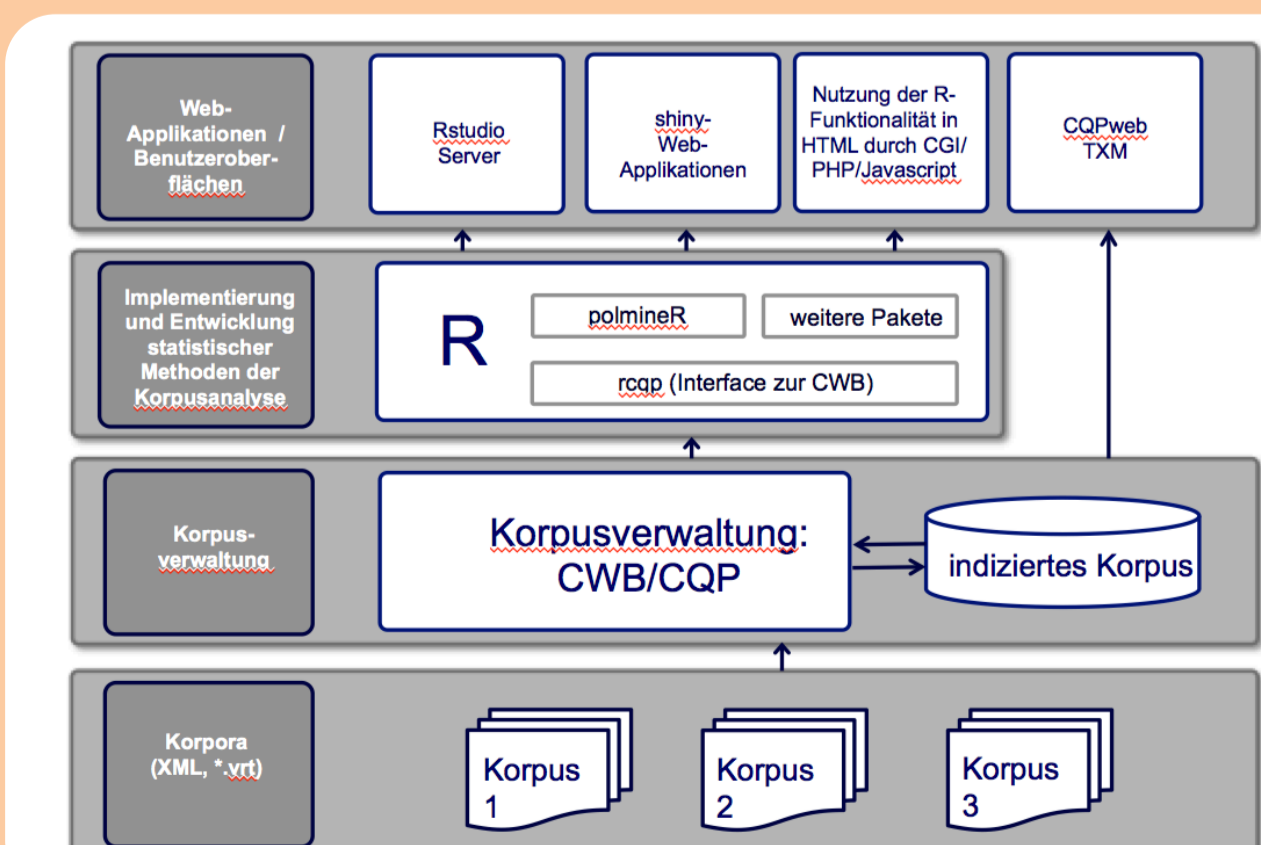
### Bereitstellung

#### Nutzungsmöglichkeiten

- Bereitstellung der Ressource über Server des IMS (Universität Stuttgart)
- Möglichkeit des Datenzugriffs über CQPweb
- Nutzung des PolMine-Toolkit (über Installation RStudio Server)
- Einbindung von Nutzern in Qualitätssicherung der Ressource
- Ziel: Kontinuierliche Verbesserung der Datenqualität

#### PolMine-Toolkit

Um die Möglichkeiten der Annotationsstruktur des Plenarprotokollkorpus voll nutzen zu können, wurde mit dem PolMine-Toolkit eine spezielle Verbindung von Corpus Workbench (CWB) und R geschaffen. Das polmineR-Paket ist das Kernstück der entsprechenden Schichtenarchitektur.



#### CWB und R

Die CWB gewährleistet Performanz und bietet mit der CQP-Syntax flexible Recherchemöglichkeiten. Das R-Paket polmineR ermöglicht die flexible Definition von Subkorpora und dient als Schnittstelle zu statistischen Auswertungsmöglichkeiten.

Funktionalität von polmineR (Auswahl):

- KWIC / Konkordanz-Analyse
- Kollokationsanalysen
- Termextraktion / Schlagwortberechnung
- Exportfunktionen (html / xlsx / Email)

siehe: [www.github.com/PolMine/polmineR](http://www.github.com/PolMine/polmineR)

#### Weitere Informationen

Website des PolMine-Projekts  
[www.polmine.de](http://www.polmine.de)

PolMine auf GitHub  
[www.github.com/PolMine](http://www.github.com/PolMine)

Internetprojekt Parl-O-Mat  
[www.parlomat.de](http://www.parlomat.de)

Kontakt  
Prof. Dr. Andreas Blätte  
Professur für Public Policy und Landespolitik  
[www.andreas-blaette.de](http://www.andreas-blaette.de)  
[andreas.blaette@uni-due.de](mailto:andreas.blaette@uni-due.de)