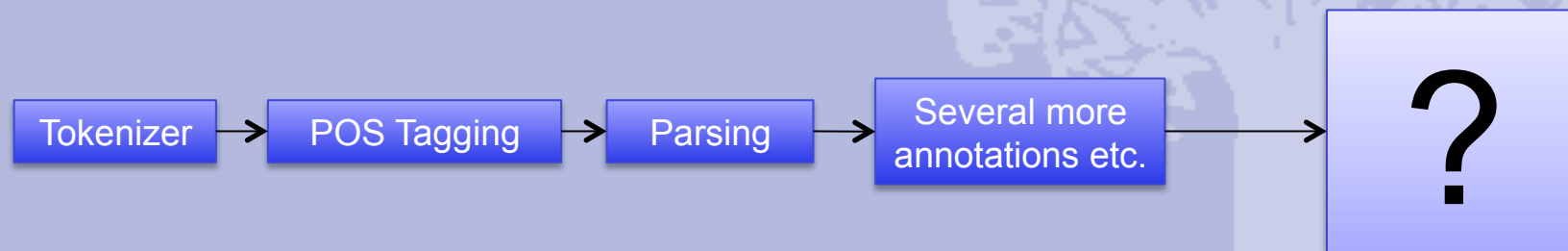


Dynamic Corpus Analyser

Motivation

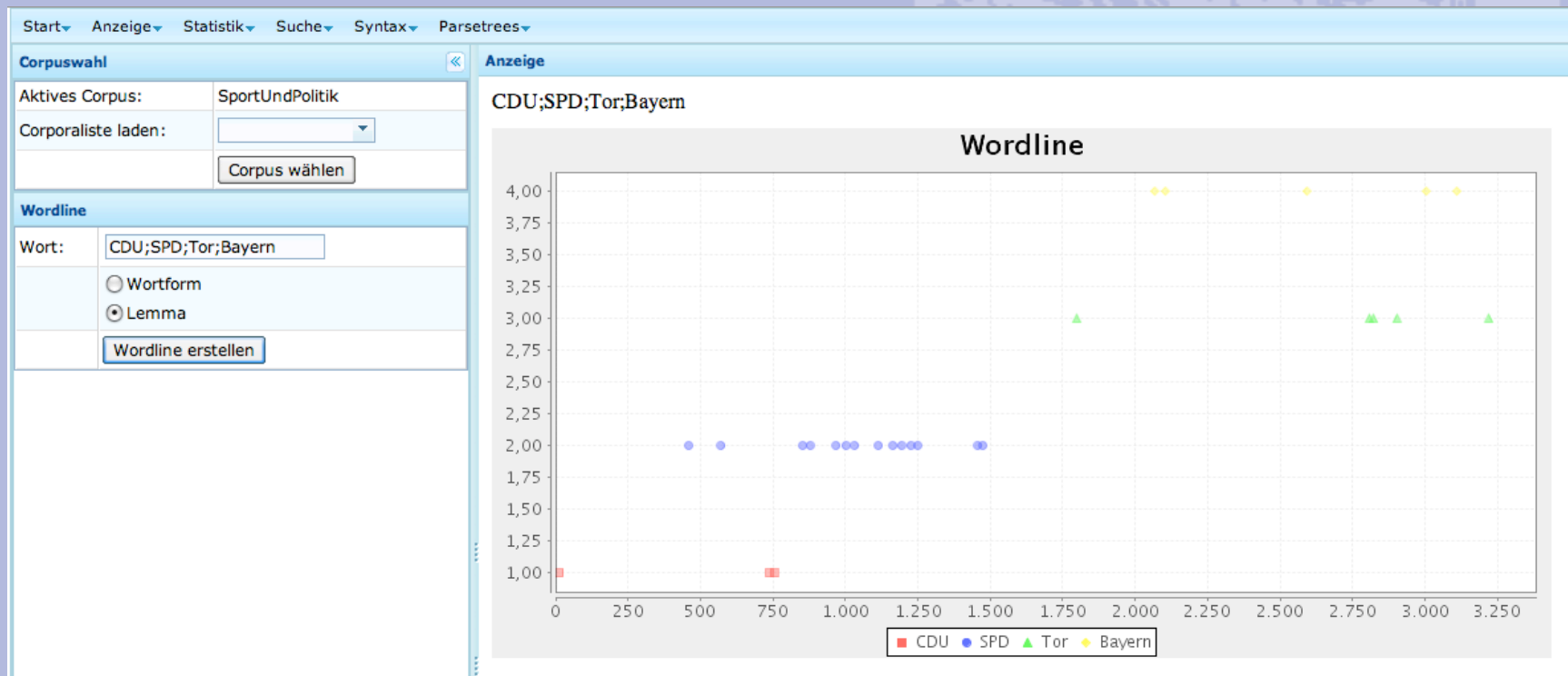
- With the D-Spin toolchain, textcorpora can be created fast and easy *on-the-fly*



- After a textcorpus is created, DCA can be used to do some (quantitative) analysis

Functionality

- Wordline – visualizes appearances of words in the linear progression of a text:



Functionality

- Some (simple) statistical analysis (entropy etc.)
- Transitional probability:

The screenshot displays the Dynamic Corpus Analyser interface. The 'Anzeige' (Display) section shows the following statistics:

- Kombination: **ART NN**
- Absolute Anzahl **ART**: 14172
- Absolute Anzahl **NN**: 30297
- Anzahl Bigramme: 10821
- Wahrscheinlichkeit, dass nach **ART** ein **NN** kommt (10821 / 14172): 76.35478408128704
- Wahrscheinlichkeit, dass vor **NN** ein **ART** kommt (10821 / 30297): 35.716407565105456

The 'Übergangswahrscheinlichkeiten' (Transition Probabilities) section shows a list of transitions with 'Abschicken' (Send) and 'POS laden' (Load POS) buttons.

Einige Beispiele (Some Examples):

Der	Schatten
des	Kardinalspalastes
einem	Tische
ein	Mann
Der	Widerschein
des	Herdes
das	Licht
die	Einsamkeit
die	Stille
der	Vorzimmer

Übergangswahrscheinlichkeiten (Transition Probabilities) Bar Chart:

Transition	Count
ART	14172
NN	30297

Functionality

- Lemma statistic:

Start ▾ Anzeige ▾ Statistik ▾ Suche ▾ Syntax ▾ Parsetrees ▾

Corpuswahl

Aktives Corpus: dumas
Corporaliste laden:

Lemma Statistik

Lemma: melden

Anzeige

Implementierungen von Lemma **melden**:

melden	25
meldete	11
meldet	14
gemeldet	7
melde	2
meldeten	1

Lemma Statistik

Legend: melden (cyan), meldeten (red), melde (blue), gemeldet (green), meldete (yellow), meldet (pink)

Functionality

- Search is possible for (bi-) grams, individual words, POS of words and whole phrases with the support of regular expressions and Leveshtein distance

The image displays three screenshots of the Dynamic Corpus Analyser interface, illustrating different search functionalities. Each screenshot shows a navigation menu at the top with options: Start, Anzeige, Statistik, Suche, Syntax, and Parse.

Left Screenshot: Wort Suche
- **Corpuswahl:** Aktives Corpus: dumas; Corporaliste laden: [dropdown]; Corpus wählen button.
- **Wort Suche:** Wort: melden; Search options: Exakt, Beinhaltet, Beginnt mit, Endet mit, Levenshtein max: 1; Abschicken button.

Middle Screenshot: Suche Wortpaar
- **Corpuswahl:** Aktives Corpus: dumas; Corporaliste laden: [dropdown]; Corpus wählen button.
- **Suche Wortpaar:** Erstes Wort: ich; Zweites Wort: du; Max. Abstand: 5; Anzahl Treffer: 10; Abschicken button.

Right Screenshot: Sätze anzeigen
- **Corpuswahl:** Aktives Corpus: dumas; Corporaliste laden: [dropdown]; Corpus wählen button.
- **Sätze anzeigen:** Phrase: auf dem Tisch; Abschicken button.

Functionality

- Syntax: Building of a grammar, searching for a given construct:

The screenshot shows the 'Dynamic Corpus Analyser' interface. On the left, there are controls for 'Corpuswahl' (Active Corpus: dumas) and 'Konstrukt suchen' (Construct: ART;NN). The main area displays a table of search results with columns for grammatical categories and corresponding words.

ART	NN	NN	APPR	ART	PRF	ADV	ADJA	NN	ART	NN	\$,	APPR	ART
Der	Schatten	Richelieus	In	einem	uns	schon	bekannten	Gemache	des	Kardinalspalastes	,	an	einem
APPR	PPER	VAFIN	ART	ADJA	\$,	APPRART	NN	ADJA	NN	\$,	PWAV	ART	ADJA
Hinter	ihm	war	ein	weiter	,	vom	Feuer	geröteter	Kamin	,	worin	die	brennenden
ART	NN	ART	NN	VVFIN	ART	ADJA	NN	PDAT	NN	APPR	ADV	\$,	KOUS
Der	Widerschein	des	Herdes	erhellte	die	prächtige	Kleidung	dieses	Träumers	von	rückwärts	,	während
KOUS	PIS	PDAT	ADJA	NN	\$,	PDAT	ADJA	NN	\$,	PROAV	PDAT	ADJA	KON
Wenn	man	diese	rote	Kleidung	,	diese	reichen	Spitzen	,	dazu	diese	blasse	und
PPER	VAFIN	APPR	ART	NN	ADV	ADV	ART	NN	ART	ADJA	NN	\$.	
es	war	in	der	Tat	nur	noch	der	Schatten	des	großen	Mannes	.	
ART	ADJA	NE	\$,	ART	ADJA	NN	ART	NN	\$,	PRELS	ADV	ADJD	KON
Das	geschwächte	Frankreich	,	die	verkannte	Autorität	des	Königs	,	die	abermals	schwach	und
PWS	ADV	ADV	PIS	KOKOM	PIS	ART	ADJA	VVFIN	\$,	KOUS	ART	ADJA	NN
Was	aber	noch	mehr	als	alles	das	Angeführte	zeigte	,	daß	die	rote	Kleidung
PDS	VAFIN	PDAT	ADJA	NN	\$,	PRELS	PRF	APPR	ART	NN	VVFIN	KON	APPR
das	war	diese	höhnische	Gesinnung	,	die	sich	auf	der	Straße	kundgab	und	durch
PDS	VAFIN	ADV	ART	ADJA	\$,	ADV	ADJA	NN	APPR	NN	\$,	PRELS	APPRART
das	war	endlich	ein	fernes	,	stets	wiederholtes	Dröhnen	von	Flintenschüssen	,	welche	zum
ART	NN	\$,	PPER	VVFIN	ADV	PTKNEG	\$,	KOUS	ADV	PDAT	NN	\$,	PRELS
Die	Schwachköpfe	,	sie	fühlen	also	nicht	,	daß	keineswegs	dieser	Welsche	,	der

Navigation: [1 / 435] [1 - 20 / 8691]

Further Enhancements

- More statistical analysis: integration of quantitative textlaws
- Integration of (semantic) dictionaries (GermaNet, OpenThesaurus)
- Technical enhancements (caching of results)
- Integration of more texts
- Comparison of two texts (Goethes Faust I vs. 2)
- → Grouping of texts into classes (comparison of Goethe vs. Schiller)