

Sprache oberflächlich

Datengeleitete Analyse komplexer Muster in Textkorpora für sozialwissenschaftliche Anwendungsbereiche

Noah Bubenhofer und Klaus Rothenhäusler



Heidelberg Center for American Studies (HCA)

22. 10. 2009 / Computational Linguistics Colloquium, Heidelberg

- 1 Einführung
- 2 Technische Infrastruktur und Umsetzung
- 3 Analyseergebnisse

Rede 1

Rede 2

Ebenen der linguistische Analyse von politischen Reden

- Rhetorik: Welche klassischen rhetorischen Stilmittel werden verwendet?
- Stilistik: Welches Register wird verwendet? Durch welche sprachlichen Phänomene zeigt sich dieses Register?
- Diskursanalyse: Warum wird in der Rede der Redegegenstand so ausgedrückt, wie er ausgedrückt wird? Welche Denk- und Argumentationsmuster verbergen sich dahinter?

Methoden der linguistische Analyse von politischen Reden

- Hermeneutik: Lesen, verstehen, Schlüsse ziehen
- Inhaltsanalyse: Coderaster aufstellen, codieren, zählen, Schlüsse ziehen
- Korpuslinguistik: These über Sprachgebrauch operationalisieren, recherchieren, auswerten, Schlüsse ziehen

→ Deduktive Verfahren der Analyse; Ausgangspunkt: Hypothesen

Projekt **semtracks**: Team



Joachim Scharloth
Saskia Vola
Klaus Rothenhäusler
Suarès Tamekue
Martin Klimke
Noah Bubenhofer
Sarah Ebling

Finanzierung: Universität Heidelberg, Innovationsfonds „Frontier“
im Rahmen der Exzellenzinitiative

Methodische Ziele des Projekts:

- Ausgangsbasis sind große Korpora schriftlicher Sprache.
- Im Fokus der Analyse steht die sprachliche Oberfläche mit verschiedenen Merkmalen:
 - Tokens
 - Lexeme
 - POS-Annotationen
 - Syntaktische Annotationen
 - n-Gramme bestehend aus Tokens/Lexemen oder einer Kombination aller möglicher Merkmale
- Verschiedene Teilkorpora werden auf die Frequenz dieser Merkmale getestet: Erstellung einer **semantischen Matrix** von typischen Merkmalen für das jeweilige Teilkorpus.

Die Wahl der Teilkorpora kann sich orientieren an:

- Textsorten (Leserbriefe vs. andere Zeitungstexte)
- Personen (Angela Merkel vs. Frank-Walter Steinmeier)
- Zeitabschnitte (Web-Diskussionsforum: Threads zum Thema USA in den Jahren 2000 bis 2009)

Dabei liegt das **Interesse der Analyse** darin, Merkmale der sprachlichen Oberfläche zu finden, die eine **semantische Ladung** aufweisen, also in bestimmten Textsorten, Zeitabschnitten oder bei bestimmten Personen kontextspezifische Bedeutung tragen.

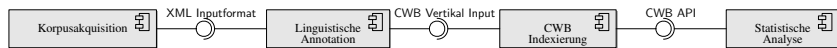
Entwicklung und Test der Methoden am Beispiel von:

- US-Präsidentschaftswahl 2008: John McCain vs. Barack Obama
- Deutschland Bundestagswahl 2009

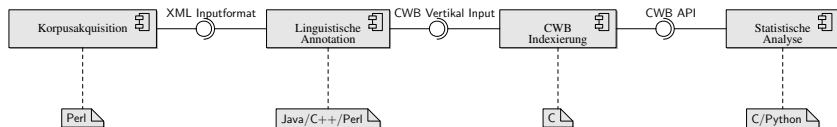


- Ressourcen:
 - ca 850 Cores (Opteronen, Xeons)
 - 4–32 GB RAM
 - Verbindung über Ethernet
- Beteiligt an ATLAS
- Kollaborationslösung: trac

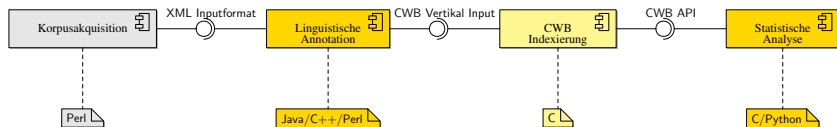
Systemarchitektur



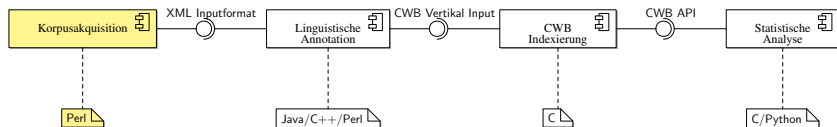
Systemarchitektur



Systemarchitektur

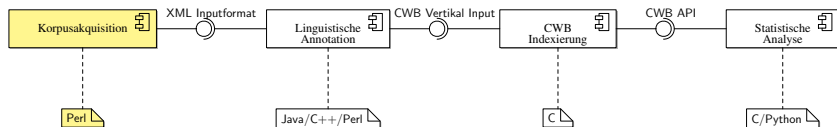


Systemarchitektur – Korpusakquisition



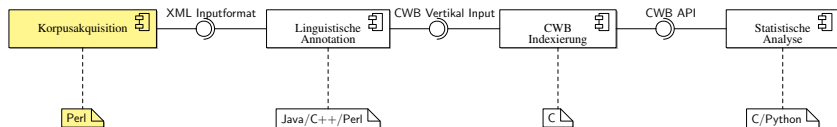
- Web basierte Korpora
- Integration geplant

Systemarchitektur – Korpusakquisition



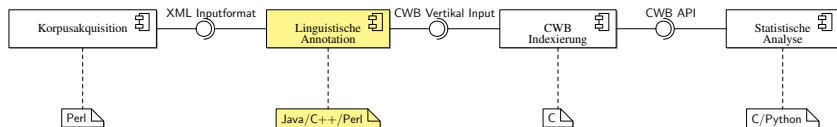
- Web basierte Korpora
- Integration geplant

Systemarchitektur – Korpusakquisition



- Web basierte Korpora
- Integration geplant

Systemarchitektur – Linguistische Analyse

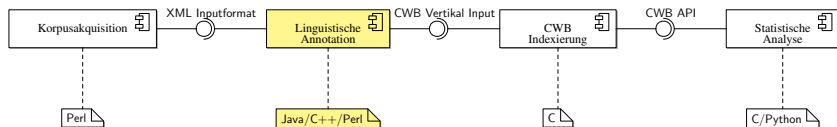


- Pipeline basierend auf

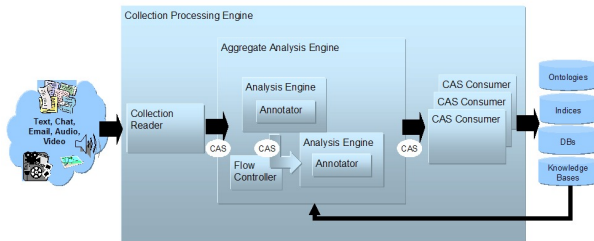


Unstructured
Information Management
Architecture
An Apache Incubator Project

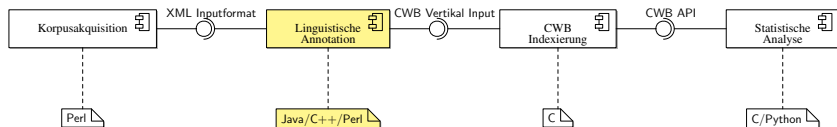
Systemarchitektur – Linguistische Analyse



- Pipeline basierend auf



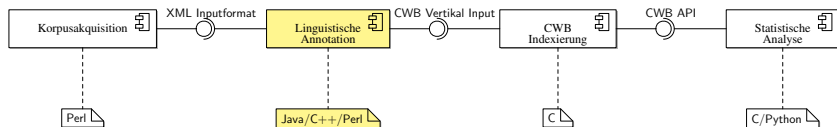
Systemarchitektur – Linguistische Analyse



Annotatoren

- Tokenisierung
- Part of Speech Tagging (TreeTagger)
- Morphologische Analyse (Morphisto), z.B. *Steigerungsformen*
- Semantische Wortklassen, z.B. *Itensivierer, Heckenausdrücke*
- Satz and Gliedsatz Annotation (Regelbasiert in Kombination mit maschinellem Lerner)
 - beinhaltet *Tempus, Modus und Aktionsart* Identifizierung

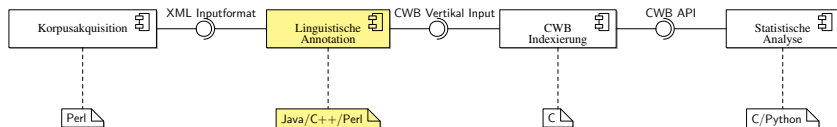
Systemarchitektur – Linguistische Analyse



Annotatoren

- Tokenisierung
- Part of Speech Tagging (TreeTagger)
- Morphologische Analyse (Morphisto), z.B. *Steigerungsformen*
- Semantische Wortklassen, z.B. *Itensivierer, Heckenausdrücke*
- Satz and Gliedsatz Annotation (Regelbasiert in Kombination mit maschinellem Lerner)
 - beinhaltet *Tempus, Modus und Aktionsart* Identifizierung

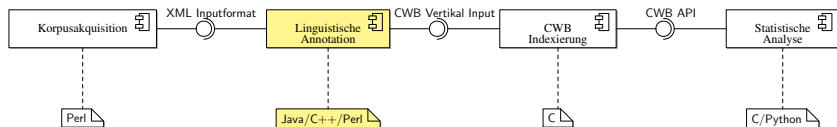
Systemarchitektur – Linguistische Analyse



Annotatoren

- Tokenisierung
- Part of Speech Tagging (TreeTagger)
- Morphologische Analyse (Morphisto), z.B. *Steigerungsformen*
- Semantische Wortklassen, z.B. *Itensivierer, Heckenausdrücke*
- Satz and Gliedsatz Annotation (Regelbasiert in Kombination mit maschinellem Lerner)
 - beinhaltet *Tempus, Modus und Aktionsart* Identifizierung

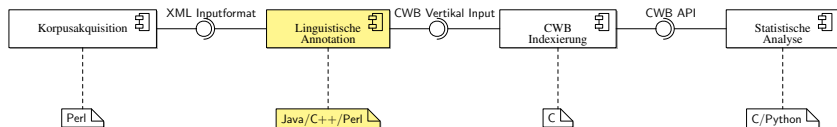
Systemarchitektur – Linguistische Analyse



Annotatoren

- Tokenisierung
- Part of Speech Tagging (TreeTagger)
- Morphologische Analyse (Morphisto), z.B. *Steigerungsformen*
- Semantische Wortklassen, z.B. *Itensivierer, Heckenausdrücke*
- Satz and Gliedsatz Annotation (Regelbasiert in Kombination mit maschinellem Lerner)
 - beinhaltet *Tempus, Modus und Aktionsart* Identifizierung

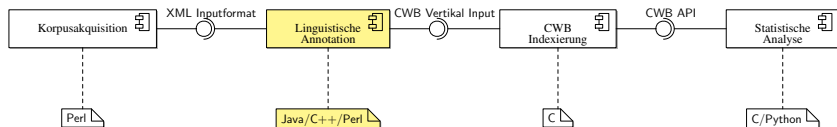
Systemarchitektur – Linguistische Analyse



Annotatoren

- Tokenisierung
- Part of Speech Tagging (TreeTagger)
- Morphologische Analyse (Morphisto), z.B. *Steigerungsformen*
- Semantische Wortklassen, z.B. *Itensivierer, Heckenausdrücke*
- Satz and Gliedsatz Annotation (Regelbasiert in Kombination mit maschinellem Lerner)
 - beinhaltet *Tempus, Modus und Aktionsart* Identifizierung

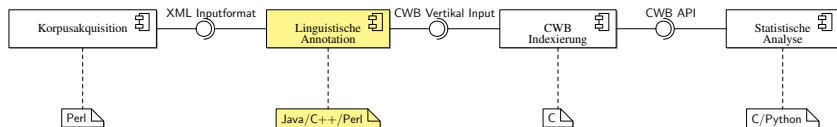
Systemarchitektur – Linguistische Analyse



Annotatoren

- Tokenisierung
- Part of Speech Tagging (TreeTagger)
- Morphologische Analyse (Morphisto), z.B. *Steigerungsformen*
- Semantische Wortklassen, z.B. *Itensivierer, Heckenausdrücke*
- Satz and Gliedsatz Annotation (Regelbasiert in Kombination mit maschinellem Lerner)
 - beinhaltet *Tempus, Modus und Aktionsart* Identifizierung

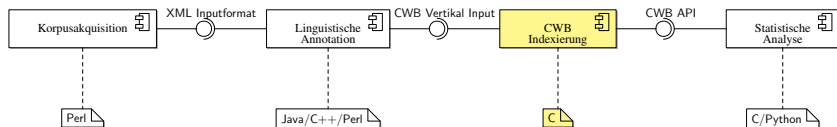
Systemarchitektur – Linguistische Analyse



Annotatoren

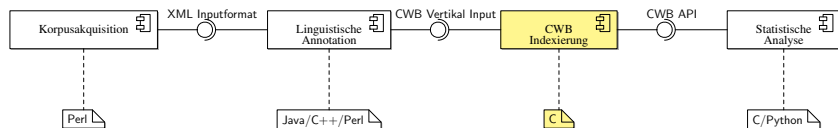
- Tokenisierung
- Part of Speech Tagging (TreeTagger)
- Morphologische Analyse (Morphisto), z.B. *Steigerungsformen*
- Semantische Wortklassen, z.B. *Itensivierer, Heckenausdrücke*
- Satz and Gliedsatz Annotation (Regelbasiert in Kombination mit maschinellem Lerner)
 - beinhaltet *Tempus, Modus* und *Aktionsart* Identifizierung

Systemarchitektur – CWB Indexierung



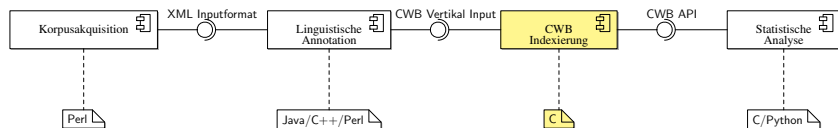
- Datensenke:
- Korpora Management
- Invertierter Index für Text und Annotationen
- Komplexe Anfragemöglichkeiten
 - Reguläre Ausdrücke auf Text, Annotationen und Attributen
 - Einschränkung des Suchbereichs auf annotierte Regionen
- Effiziente Implementierung von statistischen Analysen

Systemarchitektur – CWB Indexierung



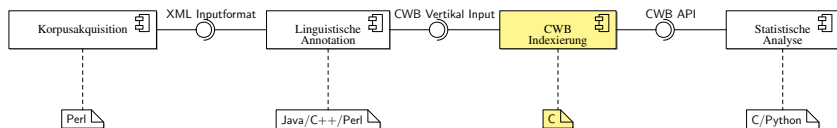
- Datensenke:
- Korpora Management
- Invertierter Index für Text und Annotationen
- Komplexe Anfragemöglichkeiten
 - Reguläre Ausdrücke auf Text, Annotationen und Attributen
 - Einschränkung des Suchbereichs auf annotierte Regionen
- Effiziente Implementierung von statistischen Analysen

Systemarchitektur – CWB Indexierung



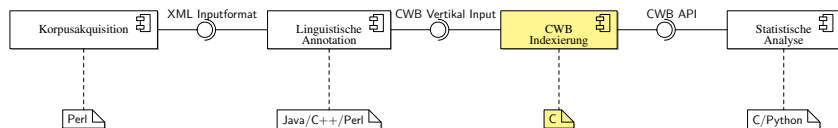
- Datensenke:
- Korpora Management
- Invertierter Index für Text und Annotationen
- Komplexe Anfragemöglichkeiten
 - Reguläre Ausdrücke auf Text, Annotationen und Attributen
 - Einschränkung des Suchbereichs auf annotierte Regionen
- Effiziente Implementierung von statistischen Analysen

Systemarchitektur – CWB Indexierung



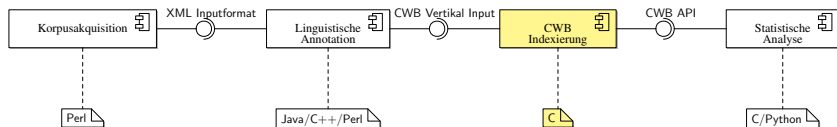
- Datensenke:
- Korpora Management
- Invertierter Index für Text und Annotationen
- Komplexe Anfragemöglichkeiten
 - Reguläre Ausdrücke auf Text, Annotationen und Attributen
 - Einschränkung des Suchbereichs auf annotierte Regionen
- Effiziente Implementierung von statistischen Analysen

Systemarchitektur – CWB Indexierung



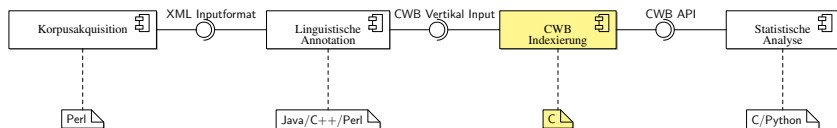
- Datensenke:
- Korpora Management
- Invertierter Index für Text und Annotationen
- Komplexe Anfragemöglichkeiten
 - Reguläre Ausdrücke auf Text, Annotationen und Attributen
 - Einschränkung des Suchbereichs auf annotierte Regionen
- Effiziente Implementierung von statistischen Analysen

Systemarchitektur – CWB Indexierung



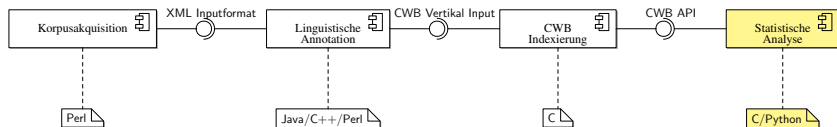
- Datensenke:
- Korpora Management
- Invertierter Index für Text und Annotationen
- Komplexe Anfragemöglichkeiten
 - Reguläre Ausdrücke auf Text, Annotationen und Attributen
 - Einschränkung des Suchbereichs auf annotierte Regionen
- Effiziente Implementierung von statistischen Analysen

Systemarchitektur – CWB Indexierung



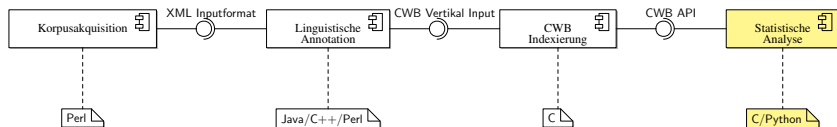
- Datensenke:
- Korpora Management
- Invertierter Index für Text und Annotationen
- Komplexe Anfragemöglichkeiten
 - Reguläre Ausdrücke auf Text, Annotationen und Attributen
 - Einschränkung des Suchbereichs auf annotierte Regionen
- Effiziente Implementierung von statistischen Analysen

Systemarchitektur – Statistische Analyse



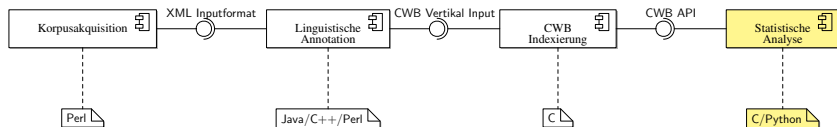
- Buzzwords (komparativ)
 - Signifikanztest auf Indexfrequenzen
- N-Gramme (komparativ)
 - nutzt Suffixarrays (Yamamoto and Church, 2001)
 - $O(N)$ Laufzeit und Speicher
- Kollokationen
 - Korpuslinguistisches Standardverfahren
 - mehrstufig mit Wortgraphen als Output
 - verschiedene Assoziationsmaße
- *Wortraummodelle zur Identifizierung unterschiedlicher Verwendungsweisen*

Systemarchitektur – Statistische Analyse



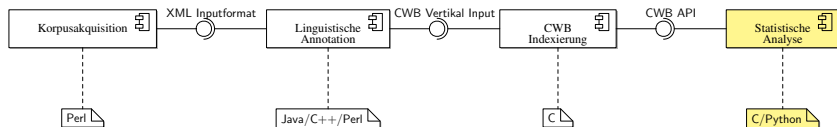
- Buzzwords (komparativ)
 - Signifikanztest auf Indexfrequenzen
- N-Gramme (komparativ)
 - nutzt Suffixarrays (Yamamoto and Church, 2001)
 - $O(N)$ Laufzeit und Speicher
- Kollokationen
 - Korpuslinguistisches Standardverfahren
 - mehrstufig mit Wortgraphen als Output
 - verschiedene Assoziationsmaße
- *Wortraummodelle zur Identifizierung unterschiedlicher Verwendungsweisen*

Systemarchitektur – Statistische Analyse



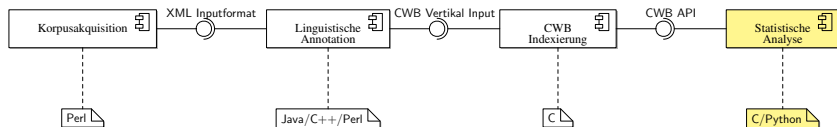
- Buzzwords (komparativ)
 - Signifikanztest auf Indexfrequenzen
- N-Gramme (komparativ)
 - nutzt Suffixarrays (Yamamoto and Church, 2001)
 - $O(N)$ Laufzeit und Speicher
- Kollokationen
 - Korpuslinguistisches Standardverfahren
 - mehrstufig mit Wortgraphen als Output
 - verschiedene Assoziationsmaße
- *Wortraummodelle zur Identifizierung unterschiedlicher Verwendungsweisen*

Systemarchitektur – Statistische Analyse



- Buzzwords (komparativ)
 - Signifikanztest auf Indexfrequenzen
- N-Gramme (komparativ)
 - nutzt Suffixarrays (Yamamoto and Church, 2001)
 - $O(N)$ Laufzeit und Speicher
- Kollokationen
 - Korpuslinguistisches Standardverfahren
 - mehrstufig mit Wortgraphen als Output
 - verschiedene Assoziationsmaße
- *Wortraummodelle zur Identifizierung unterschiedlicher Verwendungsweisen*

Systemarchitektur – Statistische Analyse



- Buzzwords (komparativ)
 - Signifikanztest auf Indexfrequenzen
- N-Gramme (komparativ)
 - nutzt Suffixarrays (Yamamoto and Church, 2001)
 - $O(N)$ Laufzeit und Speicher
- Kollokationen
 - Korpuslinguistisches Standardverfahren
 - mehrstufig mit Wortgraphen als Output
 - verschiedene Assoziationsmaße
- *Wortraummodelle zur Identifizierung unterschiedlicher Verwendungsweisen*



U.S. Presidential Campaign '08: A Semantic Matrix Analysis

Latest Entries:

- Visions of America
- Sarah Palin - Between Alaska, Joe Sixpack and the White House - Part II
- Sarah Palin - Between Alaska, Joe Sixpack and the White House - Part I
- The Final Stretch: Consistency and the Hope for Game Changers

About:

- semtracks
- Introduction
- Authors
- Glossary
- Terms and Conditions
-  RSS-Feed

Introduction

Language is a core medium for the social construction of today's reality. In political rhetoric, language is used to influence the voters' way of thinking and to shape their emotions and attitudes. But language is not only a medium of persuasion and manipulation. It also reveals the speaker's underlying values and attitudes and tells us something about the Zeitgeist.

The current U.S. presidential campaign is one of the major events that illustrates how the specific semantics of a candidate's rhetoric are profiled, debated and contested in the public arena.

SEMTRACKS Political Tracker provides regular analyses of the campaign speeches by the U.S. presidential candidates in 2008, John McCain and Barack Obama. Based on our semantic search and tracking technology, we compare the candidates' language and the characteristic features of their political rhetoric.

SEMTRACKS Political Tracker does not just count words and provides you with frequency lists, like others do. Rather, we focus on the significant differences between the ways politicians use language.

In all our analyses, the margin of error is max. 0.05%; in other words, with a probability of 95% and higher the differences between the candidate's language in his campaign speeches are not random but are the result of specific linguistic strategies in the creation of political messages.

SEMTRACKS thus tracks the unique linguistic features of a candidate that contribute to the overall semantic matrices their campaigns try to create.

Our blog starts with the analysis of simple linguistic categories and will gradually provide the reader with more complex examples.

For a German summary of our latest research results, see here: "Die 'Wort-Wahl': Barack Obama und John McCains rhetorischer Kampf ums Weiße Haus." (Oct 25, 2008)

Tracking the Candidates:

[Biden vs. Palin](#)

[McCain](#)

[McCain vs. Obama](#)

[Obama](#)

[Palin](#)

Tracking through Time:

[Covering Various Phases](#)

[Debate 1](#)

[Debate 2](#)

[Debate 3](#)

[The Final Weeks](#)

[VP Debate](#)

Linguistic Categories:

<http://semtracks.com/politicaltracker/>

Dynamisches Korpus:

- TV-Debatten McCain/Obama: 48.000 Tokens
- (Wahlkampf-)Reden 2002/2007–2008: 1,1 Mio. Tokens

Analysedimensionen:

- Debatten-/Reden-Analysen: Kontrastierung McCain/Obama
- Diachrone Analysen: Entwicklung des Wahlkampfs bei McCain und Obama
- Thematische Analysen: Diachrone oder synchrone Differenzen im Sprachgebrauch zu bestimmten Themen (Wirtschaft, Irak, Terrorismus etc.)

Ergebnisse: US-Wahlen

Wortwolken: Zweite TV-Debatte

McCain

again. America America's
Americans certainly fight fine fix
friend his I'll increase know
leave like **my** never
Obama Obama's
obviously once organization other proposal raise really
record reform Senate strategy taxis
thank them together too tough value vote we'll

© semtracks

Obama

actually bin budget **can't** care change class
cut deal energy enormous example go happen
help her important Iran it just Laden
last **make**
McCain
McCain's mean more need next president
provide set **so** something sure system that's
us We're **we've** year you've your

semtracks
www.semtracks.com

Ergebnisse: US-Wahlen

Wortwolken: Dritte TV-Debatte

McCain

America angry around country course floor
government he him home I'm increase Iraq Joe know maybe nuclear

Obama

Obama's party raise remark Senator she She's spending spread
take tax They're United vote wealth world you

Obama

additional behalf break car crisis economic employer-based
example financial focus get go happen here here's important
insurance invest last less little make McCain
McCain's policy president provide serious solve support term
think though us we're work

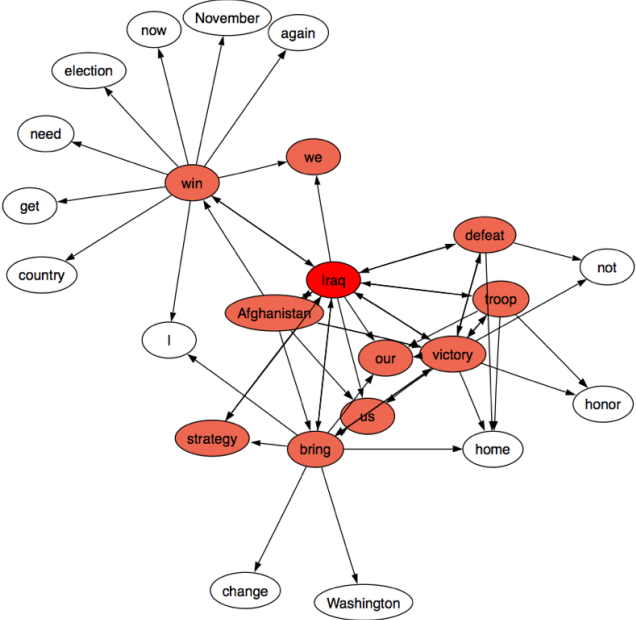
© semtracks

semtracks

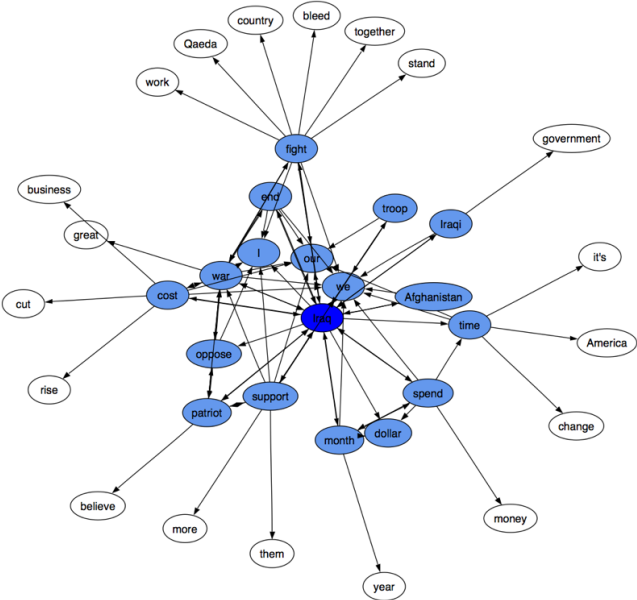
Ergebnisse: US-Wahlen

Pronoun	Significance Level	Relative Frequency Factor
McCain		
my	< 0.0001	1,66
their	< 0.0001	1,19
he	0,00016	1,26
I	0,00776	1,08
Obama		
we	< 0.0001	1,4
you	< 0.0001	1,5
us	< 0.0001	1,33
yourself	0,00028	6,17
they	0,03107	1,1

Ergebnisse: US-Wahlen



Ergebnisse: US-Wahlen



Ergebnisse: Bundestagswahl

semtracks

political tracker – Bundestagswahl '09
Eine Semantische Matrixanalyse

Analysen

- » Koalitionsoptionen aus tiefensemantischer Perspektive
- » Rhetorik des Populismus? Die Reden Guido Westerwelles und Oskar Lafontaine im Vergleich
- » Kandidenduelle auf Distanz
- » Stammwähler: Strategiewechsel – Merkel: rhetorische Kontinuität.
- » Wie nah stehen sich Schwarz und Grün?

Rubriken

- » Kontakt
- » Leistungskatalog
- » Nutzungsbedingungen
- » Team political tracker Bundestagswahlen
- » Über "semtracks"
- » Wahlprogramme auf einen Blick

Suchen

Über semtracks

- » RSS-Feed Artikel
- » Presse
- » political tracker U.S. Presidential Campaign '08

- » Anmelden

Über "semtracks"

Wie unterscheiden sich SPD, LINKE und GRÜNE in ihrer Wortwahl? Wie nahe stehen sich CDU und GRÜNE aus sprachlicher Sicht? Wie unterscheidet sich die politische Rhetorik von Angela Merkel und Frank-Walter Steinmeier? Mit welchen Schlagwörtern werben die Parteien um Wähler? Hat Angela Merkel oder Frank-Walter Steinmeier das Fernsehduell gewonnen?

Die Forschergruppe "semtracks" liefert mit ihren linguistischen Analysen Antworten auf diese und weitere Fragen rund um die Brennpunkte der deutschen Politik. Bis zur Bundestagswahl am 27. September 2009 publiziert "semtracks" auf dieser Plattform in regelmäßigen Abständen semantische Analysen. Eine Übersicht über die Themen findet sich auf dieser Seite.

"semtracks" führt auch Analysen auf Auftragsbasis durch. Unsere Leistungen im Detail finden Sie hier: [Leistungskatalog](#), [Leistungskatalog \(PDF\)](#)

Wahlprogramme auf einen Blick



<http://semtracks.com/politicaltracker/bundestagswahl/>

semtracks

Ergebnisse: Bundestagswahl

Korpus

190 Reden
(330.084
Tokens)



278 Reden
(768.549
Tokens)

Merkel: Relativierung: Heckenausdrücke (1/2)

- **. PPER VVFIN , dass**
 - . Ich glaube , dass
 - . Wir sehen , dass
 - . Sie wissen , dass
 - . Wir wissen , dass
 - . Ich weiß , dass
- **Ich VVFIN , PPER VAFIN**
 - Ich glaube , es wird
 - Ich denke , wir werden
 - Ich denke , wir sind
 - Ich glaube , es ist
 - Ich finde , wir sind

- . **PPER VVFIN , PPER**
 - . Ich glaube , ich
 - . Ich hoffe , ihn
 - . Ich glaube , es
 - . Ich finde , wir
 - . Ich denke , wir

- **\$.** **Aber ich VVFIN** ,
 - . Aber ich vermute ,
 - . Aber ich finde ,
 - . Aber ich glaube ,
 - . Aber ich hoffe ,
 - . Aber ich denke ,
- **\$.** **Aber PPER VMFIN ADV**
 - . Aber wir wollen auch
 - . Aber wir dürfen auch
 - . Aber wir müssen natürlich
 - . Aber ich möchte auch
 - . Aber ich möchte vielleicht

- **Aber ich VVFIN PPER ADV**

Aber ich stimme Ihnen auch

Aber ich vertrete sie hier

Aber ich sage Ihnen hier

Aber ich sage Ihnen ganz

Aber ich bitte Sie auch

- **\$. Deshalb VAFIN es ADV**
 - . Deshalb ist es auch
 - . Deshalb ist es ganz
 - . Deshalb ist es beispielsweise
 - . Deshalb ist es heute
 - . Deshalb ist es so
- **. Deshalb VMFIN PPER ADV**
 - . Deshalb müssen wir wieder
 - . Deshalb möchte ich noch
 - . Deshalb müssen wir auch
 - . Deshalb darf es natürlich
 - . Deshalb möchte ich hier

- . **Deshalb VAFIN ART NN**
 - . Deshalb ist die Bundesregierung
 - . Deshalb hat der Bundespräsident
 - . Deshalb ist die Haushaltskonsolidierung
 - . Deshalb ist die Zahl
 - . Deshalb wird das Thema

Merkel: Emotionalität: Emphase (1/2)

- **PPER VAFIN ADV ADJD ,**
Mir ist aber wichtig ,
Ich bin sehr froh ,
Ich bin auch froh ,
Wir sind auch froh ,
Ich bin zutiefst überzeugt ,
- **ART ADJD ADJA NN .**
eine vollkommen andere Situation .
ein ganz wesentlicher Weg .
die ganz große Herausforderung .
ein ganz gewichtiger Punkt .
einer ganz neuen Bedrohung .

- **PROAV VAFIN es ADV ADJD**

deshalb ist es so wichtig

Deshalb ist es ganz wichtig

Deshalb ist es so notwendig

Deshalb ist es so wichtig

Deshalb ist es so erfreulich

Merkel: Emotionalität: Intensivierende Partikel (1/2)

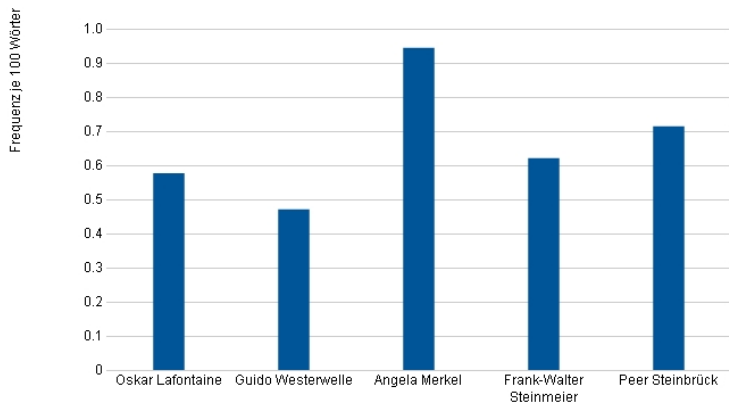


Chart by semtracks [Terms of Use](#)

Politiker

Merkel: Emotionalität: Intensivierende Partikel (2/2)

Lemma	Frequenz Steinmeier	Frequenz Merkel	Chi-Quadrat	Signifikanzniveau	x mal häufiger als bei Steinmeier
vollkommen	1	98	39,712153	< 0.0001	42,09
unglaublich	5	168	60,703953	< 0.0001	14,43
allergrößt	4	98	33,121448	< 0.0001	10,52
bestimmt	19	296	86,446104	< 0.0001	6,69
unendlich	4	40	9,192124	0,00201	4,29
absolut	5	49	11,100800	0,00064	4,20
wirklich	63	582	126,256633	< 0.0001	3,96
außerordentlich	10	81	15,723470	< 0.0001	3,47
völlig	29	213	37,568863	< 0.0001	3,15
herausragend	8	52	7,973011	0,00419	2,79
einfach	95	592	86,013080	< 0.0001	2,67
riesig	16	76	7,009167	0,00747	2,04

- **NN ART NN ART NE**

Autorität des Gouverneursrates der IAEO

Erklärung des Botschafters des Irans

Klärung des Status des Kosovo

Folge des Zerfalls der Sowjetunion

Tagen der Entscheidung der EU

- **APPR ART NN APPR NE**

nach den Gesprächen in Washington

an der Grenzschutzmission in Rafah

nach den Referenden in Frankreich

wegen des Informationsaustausches mit Großbritannien

- **ADJA NN KON ADJA NN**

sozialer Marktwirtschaft und nationaler Selbstbestimmung
andere Nachbarn und regionale Großmächte
kulturelle Identität und kulturelle Differenz
absolute Sicherheit und objektive Garantien
bedeutende Verlage und etablierte Schriftsteller

- **ADJA KON ADJA NN \$.**

arbeitsame und erfolgreiche Woche !
prominente und verdiente Europäer .
gesellschaftlichem und kulturellem Austausch .
nachrichtendienstlicher und bundespolizeilicher Ebene .
enge und partnerschaftliche Beziehungen .

- **NN , NN KON NN**

Sicherheit , Stabilität und Stärkung

Verhalten , Augenmaß und Vernunft

Gewicht , Stimme und Handlungsfähigkeit

Fragen , Frieden und Integration

Demokratie , Menschenrechte und Rechtsstaatlichkeit

- **\$. Und PPER VAFIN ADJD**

- . Und ich bin sicher
- . Und wir werden sicher
- . Und ich bin froh
- . Und ich habe aufmerksam
- . Und sie ist reich

- **NN \$. Und PPER VMFIN**

- Vereinbarungen . Und wir müssen
- Seiten . Und wir sollten
- Rolle . Und ich darf
- Kinder . Und wir wollen
- Ansatzes . Und ich will

- **\$. Und PPER VVFIN PPER**
 - . Und ich sage es
 - . Und ich verspreche Ihnen
 - . Und ich wünsche mir
 - . Und ich ermuntere Sie
 - . Und wir brauchen ihn

Guido Westerwelle und Oskar Lafontaine



Korpus:

Westerwelle 80 Reden (2003–2009), 154.298 Tokens

Lafontaine 65 Reden (2005–2009), 108.265 Tokens

Guido Westerwelle und Oskar Lafontaine: Skandalisierende Adjektive

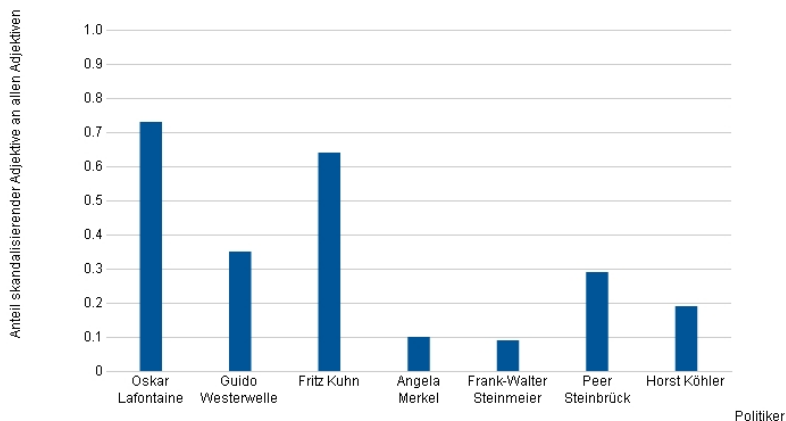


Chart by semtracks Terms of Use

Guido Westerwelle und Oskar Lafontaine: Skandalisierende Adjektive

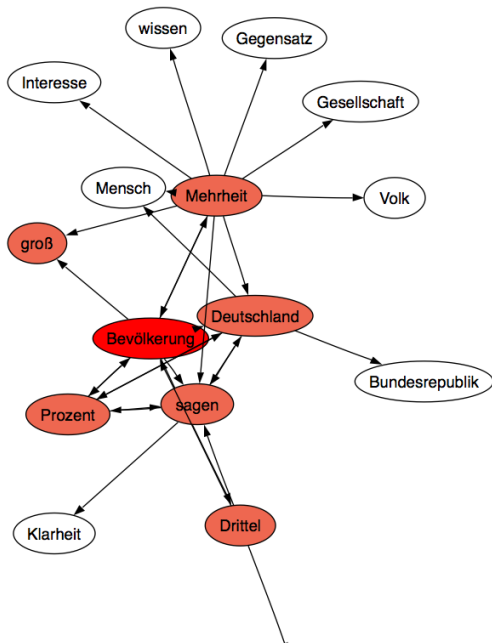
Guido Westerwelle

absurd	7
peinlich	6
billig	5
unverschämt	3
fatal	3
erschreckend	3
beispiellos	3
skandalös	2
dumm	2

Oskar Lafontaine

dumm	7
bodenlos	7
verantwortungslos	6
pervers	5
absurd	5
unverschämt	4
feige	4
billig	4
beängstigend	3
ungeheuerlich/ungeheuer	3
dreist	2

Oskar Lafontaine: *Die Mehrheit der Bevölkerung*



- Merkel: Unauffällige aber erfolgreiche Rhetorik
 - emotional
 - vorsichtig-relativierend
 - argumentativ
- Steinmeier: Nur langsamer Wandel von der schwerfälligen Außenminister-Rhetorik zum Wahlkämpfer
 - Nominalstil
 - wenig emotional
 - Stilmittel Reihungen
- Westerwelle und Lafontaine: Typische Rhetorik von Oppositionspolitikern mit feinen Differenzen:
 - Vertretung von Mehrheiten behaupten
 - skandalisierende Sprache
 - unterschiedliche Konzeptualisierungen von zentralen Begriffen wie *Staat*, *Freiheit*, *sozial* etc.

Methodisches Fazit

- Versuch, (polito-)linguistische, sozial- und kulturwissenschaftliche Thesen für empirische Analysen zu operationalisieren:
Nutzung von korpus- und computerlinguistischen Methoden.
- Ausgangsbasis ist die sprachliche Oberfläche:
Das Messen von typischen Wortkonfigurationen erlaubt es, die semantische Spezifik von Rednern/Rednerinnen, Parteien etc. herauszuarbeiten.
- Nutzung von datengeleiteten und hypothesengeleiteten Verfahren:
Mit induktiven Zugängen gewinnt man neue Ideen für zu prüfende Hypothesen.

semtracks @ Heidelberg Center for American Studies:

www.semtracks.com

bubenhofer|rothenhaeusler@semtracks.com

Yamamoto, M. and Church, K. W. (2001). Using suffix arrays to compute term frequency and document frequency for all substrings in a corpus. *Comput. Linguist.*, 27(1):1–30.