
Projects involved

- Archive of General Reference Corpora of Contemporary Written German
- Methods of Corpus Analysis and Corpus Mining

Researchers involved

- Dipl.-Ing. Cyril Belica
- Dr. Marc Kupietz
- Dipl.-Inf. Rainer Perkuhn
- Dr. Andreas Witt



Contact

Dipl.-Ing. Cyril Belica
Corpus Linguistics Programme Area
IDS – Institute for the German Language
Postfach 10 16 21
D-68016 Mannheim
Germany
Phone: +49-621-1581-0
Fax: +49-621-1581-200
Email: corpuslinguistics@ids-mannheim.de

 INSTITUT FÜR
DEUTSCHE SPRACHE

R 5, 6-13

D-68161 Mannheim

www.ids-mannheim.de

Mitglied der

Leibniz-Gemeinschaft

The Institute for the German Language (IDS) is the central institution for the study and documentation of the contemporary usage and recent history of the German language. Together with 85 other non-university research institutes and service facilities, it belongs to the Leibniz Association, one of the four major research organisations in Germany.

 INSTITUT FÜR
DEUTSCHE SPRACHE

Corpus Linguistics Programme Area

CCDB
***A Corpus-Linguistic
Research and Development
Workbench***

Within a strictly corpus-driven paradigm, an in-depth profiling of many linguistic phenomena requires fast access to massive amounts of data derived from very large corpora. The collocation data base CCDB is a corpus-linguistic research and development workbench, which serves as a *transparent lab* for the study, development, and evaluation of a methodology of data-driven exploration and modeling of language use.

The current CCDB version

- serves internally as an empirical foundation and driving force for formulating novel usage-based linguistic hypotheses
- integrates *early beta* implementations of data-driven methods currently developed: e.g., the modeling of semantic similarities, the visualization of usage aspects, and contrasting near-synonyms
- is based on a 2.2 billion words subset of the German Reference Corpus (DEREKO)
- comprises rich collocation profiles, including syntagmatic patterns and other related information, for approx. 220,000 node words
- contains for each node word up to five different collocation profiles, with up to 100,000 concordance lines each
- provides a pre-computed network of similarity relations between collocation profiles to be exploited by the data-driven methods
- can be accessed at <http://corpora.ids-mannheim.de/ccdb/>

Analysewort: **vergewissern**, Analysetyp 0

+3	5	355	wirklich auch	21	71%	sich zu vergewissern ob daß dass ... auch wir
+3	5	355	wirklich	63	80%	sich zu vergewissern ob daß dass ... auch wir
+4	-2	289	sollte vorher auch	1	100%	sollte ... vorher vergewissern ... auch
+4	-2	289	sollte vorher	9	88%	sollte [man sich] vorher [...] vergewissern ob ...
+4	-2	289	sollte auch	15	86%	sollte man sich ... vergewissern daß dass ob ...
+4	-2	289	sollte Fahrzeug	3	66%	sollte man sich vergewissern ... das Fahrzeug
+4	-2	289	sollte	71	92%	sollte man sich ... vergewissern ob daß der ...
+3	-2	271	eigenen Existenz	6	83%	der eigenen Existenz [zu] vergewissern
+3	-2	271	eigenen Geschichte	4	100%	seiner eigenen Geschichte [zu] vergewissern
+3	-2	271	eigenen Augen	3	100%	mit eigenen Augen [zu] vergewissern
+3	-2	271	eigenen	66	77%	sich der eigenen ... zu vergewissern und

Modelling Semantic Proximity, version 0.21, Hierarchical Clustering

↑ Weichsel ↓
click on arrows to expand/collapse

- Kirsche Birne Pfirsich Erdbeere Aprikose Apfel Weintraube Toma
- Landwehrkanal
- Neißer Wolga Grenzfluß Ostpreuße Galizien Flüsschen Landstrich Norden Litauen Baltikum Partisane Hügel Bergregion
- Elbe Jangtse Nebenfluß Fluss Fluß Oberlauf Hochwasser Wasse

Textwörter passend zum Suchmuster **triftig** treten als Kookkurrenzpartner bei folgenden Analysewörtern und -typen auf

Wort	0 1 2 3 4
Grundtriftigen	
Grundtriftige	
Grundtriftiger	
Argumenttriftige	
Entschuldigungstriftiger	
Scheidungsgrundtriftige	
Asylgrundtriftigen	
Gegenargumenttriftige	
vorliegentriftige	
fernbleibentriftigen	
vorbringentriftige	
anführentriftige	
fernbleibentriftiger	
Vorliegentriftiger	
vorbringentriftigen	

Folgende verwandte Kookkurrenzprofile zu **Charakteristikum** wurden gefunden (anklickbar, absteigend nach Verwandtschaftsgrad sortiert):

Merkmal
Eigenheit
Ausprägung
Element
Eigenart
Parameter
Eckpfeiler
Charakterzug
Eigenschaft
Ausdrucksform
Bezugspunkt
Ausformung
Anliegen
Stilmittel
Charakteristik
Besonderheit
Charakter
Kriterium
Ausdrucksmittel
Affinität

Modelling Semantic Proximity, version 0.19, Self-Organizing Map (initial tau: 0.02, distance: uncentered Pearson, iterations: 10)

Identität

click on a cell to toggle label locking

Kulturerbe Institution Hegemonie nationalstaatlich supranational Abgrenzung Staatlichkeit territorial	Rechte Zusammenhalt Angelegenheit Machtstruktur Struktur Errungenschaft Friedensordnung Einigungsprozeß	Eigenständigkeit Souveränität Integrität Handlungsfähigkeit Autonomie Besitzstand Bedeutung Entscheidungskompetenz	Selbständigkeit Kompetenz Gleichgewicht Selbstständigkeit Verantwortlichkeit Lebensgrundlage Glaubwürdigkeit Existenzgrundlage	Todesursache Tathergang Unschuld Staatsangehörigkeit Intimsphäre Vaterschaft Aufenthaltort Identifizierung
Selbstbestimmung national herausbilden Gesinnung Vereinnahmung Repräsentation Pluralismus Schicksalsgemeinschaft	Erbe staatsbürgerlich Unterordnung	Selbstachtung Freiheit Zusammengehörigkeitsg	Würde unwiederbringlich	Individualität Intimität Schuld Authentizität Schicksal Rückschluß Geborgenheit Lebenssinn
Zugehörigkeit ethnisch rassisch Chauvinismus Minderheit Zionismus Überfremdung Bewahrung	Selbstverständnis Selbstbehauptung Heimstatt dörflich	Zusammengehörigkeit Emanzipation kulturell Bewußtsein Selbstverwirklichung Pluralität Liberalität Urbanität	Orientierung Einzigartigkeit Identifikation Eindeutigkeit Einmaligkeit Ursprünglichkeit Leben Lebendigkeit	Potenz Profil Befriedigung Gealtungskr
religios Assimilation Entwurzelung Wiedergeburt Auslöschung Selbstvergewisserung völkisch Unterwerfung	Rückbesinnung Identitätsfindung Verkörperung	Verwurzelung Verschiedenheit Gemeinsamkeit Selbstbild Überbau Eigensinn	Bindung Bezugspunkt Homogenität Betätigung Zerrissenheit Subjektivität	Charakter Persönlichkeit Perspektive Dimension Bezug Gedächtnis Umfeld Psyche
Herkunft Tradition Religion Weltanschauung Musikkultur Religiosität Kultur Schmelztiegel	Männlichkeit Menschsein geschlechtlich Weiblichkeit Verstümmelung Fremdheit Sexualität Körperlichkeit	Vielfalt Ausdrucksform Geschlecht spezifisch Lebensform Lebenswelt genuin Vorstellungswelt	Sozialisation Ausprägung Prägung Lebensentwurf Betonung Heterogenität Empfindung widerspiegeln	Eigenheit Eigenart Besonderheit Präferenz Befindlichkeit Zuordnung Rücksicht Ausdruckswei

Printed: 04/12