





Rainer Perkuhn, Programmbereich Korpuslinguistik, IDS Mannheim



DIE GEGENWARTSSPRACHLICHEN
RESSOURCEN DES IDS UND IHRE NUTZUNG
– COSMAS & CO. –

Arbeitstreffen "Historische Lexikographie", Bullay, 16.4.2016





DAS INSTITUT FÜR DEUTSCHE SPRACHE

- ist die zentrale außeruniversitäre Einrichtung zur Erforschung und Dokumentation der deutschen Sprache in ihrem gegenwärtigen Gebrauch und in ihrer neueren Geschichte
- besteht seit 1964 in Mannheim
- wird IDS je zur Hälfte vom Bund und dem Land Baden-Württemberg (unter Beteiligung aller Länder) finanziert
- Mitglied der Leibniz-Gemeinschaft
- 210 Mitarbeiterinnen und Mitarbeiter





INSTITUT FÜR DEUTSCHE SPRACHE

Inhaltlich gegliedert in Programmbereiche, u.a.

- Programmbereich Beschreibung und Erschließung grammatischen Wissens
- Programmbereich Pragmatik und gesprochene Sprache
- Programmbereich Lexikologie und Lexikografie
- Programmbereich Korpuslinguistik
 (C. Belica, M. Kupietz, H. Lüngen, R. Perkuhn, früher auch: Holger Keibel)
 - Ausbau und Pflege der Korpora geschriebener Sprache
 - Methoden der Korpusanalyse und -erschließung



MOTIVATION

- Auftrag des IDS (aus der Satzung)
 - "... die deutsche Sprache in ihrem gegenwärtigen Gebrauch zu erforschen und zu dokumentieren ..."
- daraus folgt
 - empirische Fundierung erforderlich
- dazu unabdingbar
 - korpuslinguistischer Ansatz
 - sehr große Korpora
 - elaborierte Korpusanalysemethoden



AUSGEWÄHLTE SCHWERPUNKTE

- Das Deutsche Referenzkorpus DEREKO / Archiv der Korpora geschriebener Gegenwartssprache (http://www.ids-mannheim.de/direktion/kl/projekte/korpora.html)
- Cosmas II Recherche- und Analysesystem u.a. für das Referenzkorpus (https://cosmas2.ids-mannheim.de/cosmas2-web/)
- CCDB Kookkurrenzdatenbank (http://corpora.ids-mannheim.de/ccdb/)
- DEREWO Wortlisten zum Deutschen Referenzkorpus (http://www.ids-mannheim.de/direktion/kl/projekte/methoden/derewo.html)
- OWID / elexiko / Neologismenwörterbuch (http://www.owid.de/)
- E-Valbu (http://hypermedia.ids-mannheim.de/evalbu/index.html)



GESCHICHTE VON DEREKO UND SEINEN VORLÄUFERN

1969	Mannheimer Korpus I veröffentlicht					
	(ca. 2 Mio. Wörter – auf Lochkarten)					
1983	Recherchesoftware REFER					
1985	Bonner Zeitungskorpus veröffentlicht					
1992	COSMAS I in Betrieb					
	(IDS-Korpora: ca. 20 Mio. Wörter)					
1999	Start des BMBF-Verbundprojekts »DEREKO«					
2004	»Korpusausbau« als Dauer-Projekt etabliert					
2008	DEREKO durchbricht die 3-Milliarden-Grenze					
2015	DEREKO durchbricht die 28-Milliarden-Grenze					

■ Siehe: Belica/Teubert: Von der linguistischen Datenverarbeitung am IDS zur »Mannheimer Schule der Korpuslinguistik« in: Institut für Deutsche Sprache (Hg.): Ansichten und Einsichten. 50 Jahre Institut für Deutsche Sprache.

Mannheim: IDS



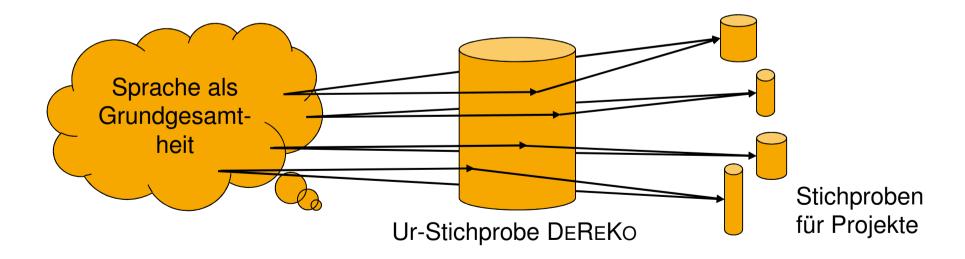
DEREKO-CHARAKTERISTIKA

- größte linguistisch motivierte Sammlung deutschsprachiger Texte
- enthält belletristische, wissenschaftliche und populärwissenschaftliche Texte, Zeitungstexte, Gebrauchstexte, verschriftlichte gesprochene Sprache ab etwa 1956, IBK
- nur vollständige Texte
- wird laufend erweitert
- Nutzung urheberrechtlich abgesichert
- im Vergleich zu historischen Korpora:
 - andere Möglichkeiten und Herausforderungen, gerade technisch, aber auch rechtlich (Menge, Vielfalt, Formate, Urheber-/Nutzungs-/Lizenz-/Persönlichkeitsrechte)



DEREKO-ZIELE

- konzipiert als »very large general purpose corpus«
- Ur-Stichprobe des gegenwärtigen Schriftsprachegebrauchs
- empirische Grundlage für die germanistische Sprachwissenschaft
- jeder Nutzer kann selbst Sub-Stichproben stratifiziert ziehen, die möglichst repräsentativ bzgl. seiner jeweiligen Fragestellung sind ("virtuelles Korpus")



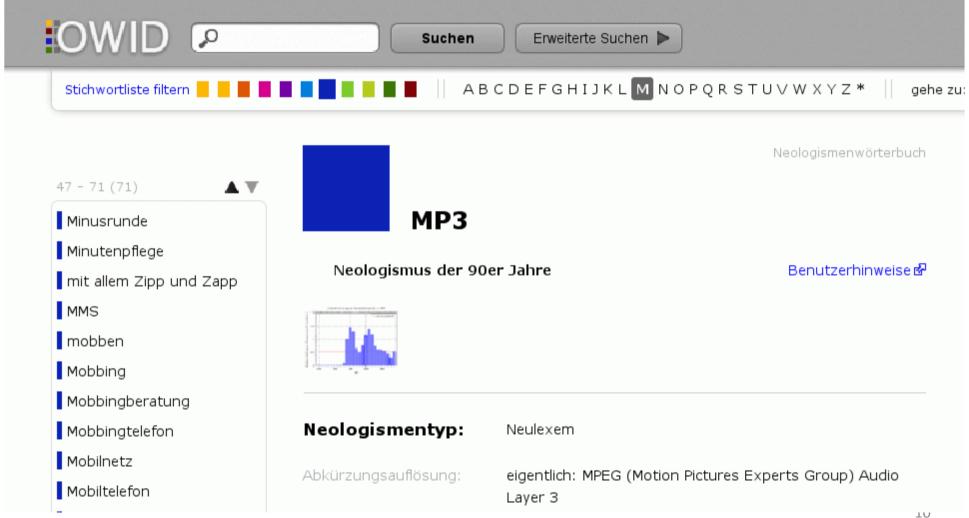


VIELEN DANK AN ALLE TEXTSPENDER!



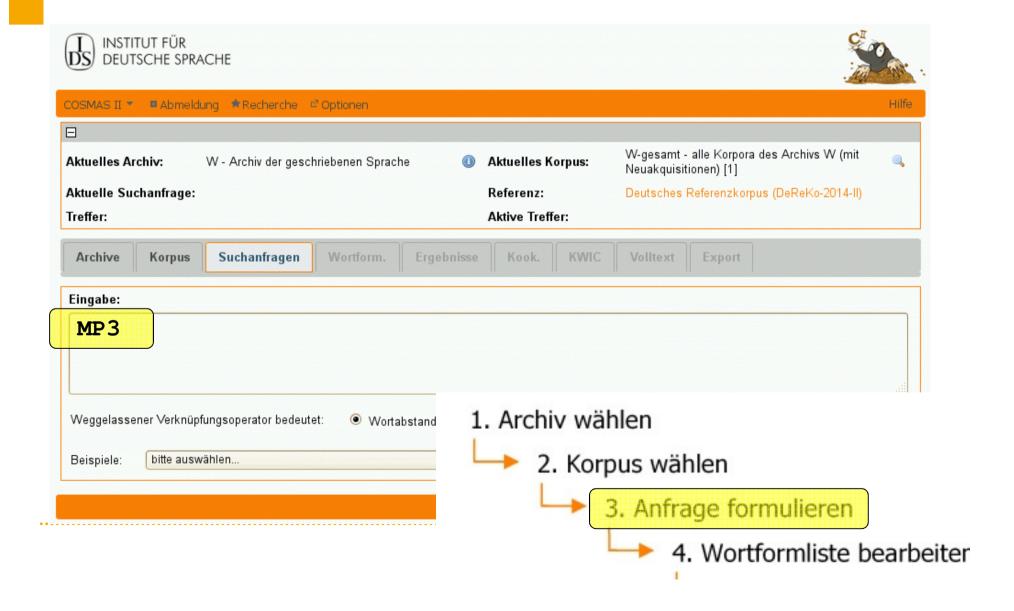


KORPUSRECHERCHE ZU GEGEBENEM LEMMA





... MITHILFE VON COSMAS II





KORPUSRECHERCHE BIS BELEG BEISPIEL: DEFINITIONSBELEG BEI NEOLOGISMEN

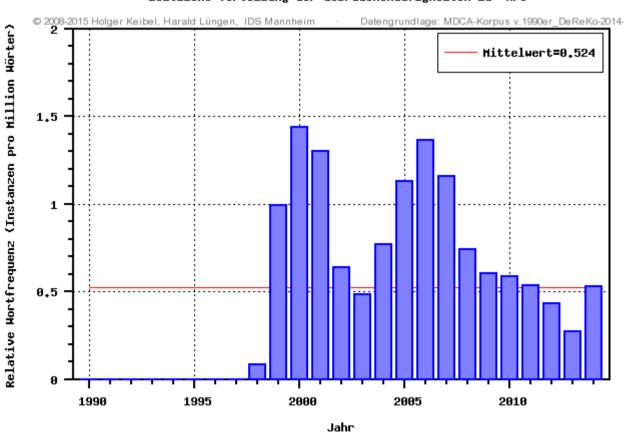
- Suche im gesamten Archiv nach "MP3"
- chronologische Sortierung der Treffermenge
- hermeneutische
 Auswertung, welcher erster
 Treffer Anforderung der
 Neologismus-Definition
 genügt





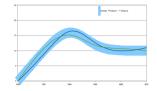
ZEITVERLAUF ALS INDIKATOR ...

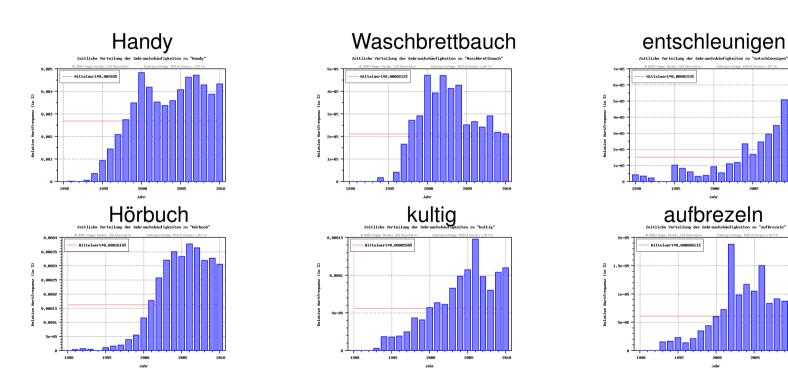
Zeitliche Verteilung der Gebrauchshäufigkeiten zu "MP3"





... FÜR NEULEXEM-KANDIDATEN





- diese Wörter haben einen Zeitverlauf, der für Neulexem notwendige Bedingung wäre
 - weitere Untersuchung, welche sind tatsächlich Neulexeme



HÄUFIGKEITEN ENTLANG VERSCHIEDENER DIMENSIONEN

- Auffälligkeit/Unauffälligkeit der Verteilung charakteristisch für verschiedene Sichten
- hohe Häufigkeit (und gleichmäßige Streuung?) guter Indikator für hohe Relevanz für
 - allgemeinsprachliche Wörterbücher?
 - Lernerwörterbücher
- vgl. auch "Kernwortschatz", niveaubezogene "Lernerwortschätze" (Nation: "mit ca. 2000 der häufigsten Wörter 85% Textüberdeckung")



DEREWO – FREQUENZBASIERTE WORTLISTEN ZUM DEUTSCHEN REFERENZKORPUS

- mehrere (kleine) Studien
 zur Erstellung von
 Häufigkeitswortlisten
 (sowohl Wort- als auch

 1.2 Der "Zentralwortschatz"
 In diesem Wörterbuch sind ca. 5500 Stichwörter in
- Dokumentation der Problemtypen insbesondere bei der Lemmatisierung (Präverbfügungen, Partizipien, Eigennamen, usw.)

Grundformen)

 letzte (große) Studie als Teilprojekt des Verbundprojekts Wechselwirkungen (DEREWO 320.000)



In diesem Wörterbuch sind ca. 5500 Stichwörter mit dem Zeichen * markiert. Dabei handelt es sich um Wörter, die in verschiedenen sogenannten Grund-

wortschätzen und Aufbauwortschätzen enthalten sind, insbesondere im Zertifikatswortschatz der Goethe-Institute. Ergänzt wurden diese Listen um Wörter, die in den Häufigkeitslisten (DeReWo) des Instituts für deutsche Sprache (IDS) und des Projekts Deutscher Wortschatz der Universität Leipzig zu den 4000 häufigsten Wörtern gehören, soweit sie für Deutschlernende wichtig erschienen. In den meisten Fällen sind dies auch Wörter, die zur Beschreibung mehrere Verwendungsangaben erfordern.

Mit dem Sternchen markierte Wörter sind also für den Lerner besonders wichtig. Dies bedeutet jedoch nicht, dass die nicht markierten Wörter "unwichtig" sind.



BEISPIEL: DUDEN 6

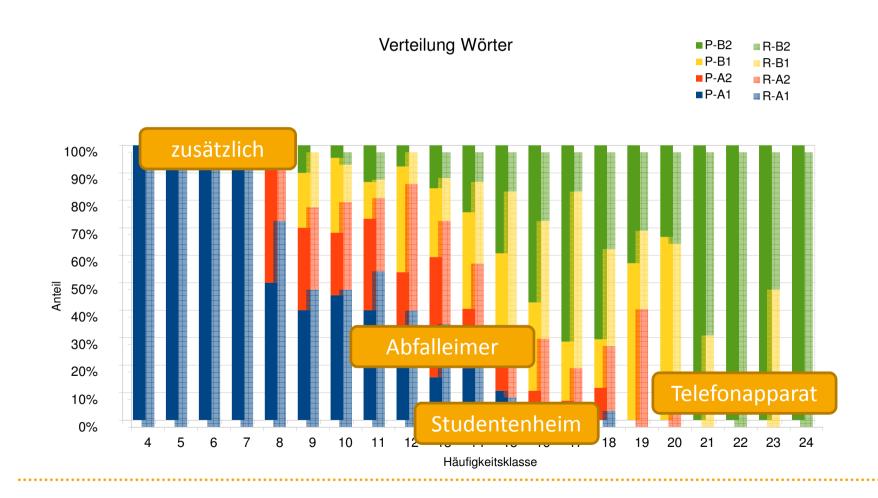
- Vorgabe: Lemmastrecke von Duden 1, 5 und 6
- Aufgabe: Vorschläge für 1.000 neue Einträge (und ggf. Streichkandidaten)
- sortiert nach Häufigkeit gemäß aktuellster DEREWO-Liste
- plus 3.000 zusätzliche häufige Einträge aus DEREWO, die in keiner der drei Listen waren
- → neue Einträge, Lemmakonzept und -gestalt



1	Lemma bzw. Sublemma	DeReWo Rang	DeReWo HK	D6	D1	D5	Lemma bzw. Sublemma	
98	beim	86	7	Х	Х	0	<	
99	was	87	7	Х	Х	0	<	
100	sehen	88	7	Х	Х	0	<	
101	deutsch	89	7	Х	0	0	<	
102	immer	90	7	Х	Х	0	<	
103	<u>doch</u>	91	7	0	Х	0	doch	
104	finden	92	7	Х	Х	0	<	
105	unter	93	7	Х	Х	0	<	
106	Prozent	94	7	Х	Х	Х	<	
107	drei	95	7	Х	Х	0	<	
108	da	96	7	Х	Х	0	<	
109	lassen	97	7	Х	Х	0	<	
110	jeder	98	7	Х	Х	0	<	
111	<u>jedes</u>	98	7	0	0	0	jedes	4
112	<u>Zeit</u>	99	7	0	Х	0	Zeit	
113	zwischen	100	7	Х	Х	0	<	
114	Kind	101	7	Х	Х	0	<	
115	hier	102	7	Х	Х	0	<	
116	seit	103	7	Х	Х	0	<	
117	bereits	104	7	Х	Х	0	<	
118	vor	105	7	Х	Х	0	<	
119	Euro	106	7	Х	Х	Х	<	
120	Tag	107	7	Х	Х	Х	<	
121	<u>beide</u>	108	7	0	Х	0	beide	
122	heute	109	7	Х	Х	0	<	
123	alt	110	7	Х	Х	0	<	
124	nun	111	7	Х	Х	0	<	
125	damit	112	7	Х	Х	0	<	
126	bleiben	113	7	Х	Х	0	<	
127	jetzt	114	7	Х	Х	0	<	
128	sehr	115	7	Х	Х	0	<	
100		110	-	10	10	-		



VERGLEICH REFERENZNIVEAU NACH PROFILE DEUTSCH MIT KORPUSHÄUFIGKEITEN





ILLUSTRATIVE ANGABEN IM WORTARTIKEL

- neben explikativen Angaben, Paraphrasen und Belegen noch viele illustrierende Angaben, die in Nennform und Komplexität variieren von kurzen Phrasen bis hin zu authentisch wirkenden ganzen Sätzen, teilweise mit Aufzählungen und Platzhalterausdrücken (jmd. oder etw.)
 - den <u>Aufzug</u> benutzen
 - Im ersten <u>Block</u> behandelt das Seminar ...
 - Er ist nicht krank, sondern er <u>macht</u> wieder einmal <u>blau</u>.
 - vor Anstrengung/Scham/Wut <u>puterrot</u> werden
 - Der Wind wirbelt Blätter/Schnee/Staub auf.
 - Ich muss noch (den Fußboden/das Bad) <u>aufwischen</u>.





N-GRAMM-ANALYSE, KOOKKURRENZANALYSE

- Verfahren zur Bestimmung von Wortkombinationen
 - Wortfolgen oder Kombinationen auf variierenden Positionen
 - statt Häufigkeit heutzutage statistische Bewertung des Ko-Vorkommens (Ko-Okkurrenz)
- vieles, was erkannt wird, kommt den illustrierenden Angaben sehr nahe (aber: Freiheiten bei Variation bei Formen und positioneller Anordnung)
- Analysen bieten riechen Fundus, wenn gewünscht wird, dass möglichst (aktuell) typische Verwendungsweisen angegeben werden



WAHRIG WÖRTERBUCH DAF (2008)

Aulge (das; -s, -n) 1 (Anat.) Sehorgan des Menschen u. der Tiere; (1) S. 146; Zus Augenfarbe, Augenlid, Augenränder 2 (Bot.) ruhende Knospe, Knospenansatz 3 Punkt, Figur, Zahl (auf Würfeln, Dominosteinen, Spielkarten usw.); Zus Augenzahl 4 (auf der Suppe, Soße) schwimmender Fetttropfen; Zus Fettauge b das ~ des Gesetzes die Polizei b aufgehen: jetzt gehen mir die ~n auf! (übertr.) jetzt erkenne ich die Zusammenhänge!; mach doch die ~n auf! (umg.) pass doch auf!; die ~n aufschlagen aufwachen; die ~n brennen, tränen mir; ihre ~n funkelten, strahlten; ein (sehr wachsames) ~ auf etwas od. jmdn. haben etwas od. jmdn. sorglich beobachten,

fen können ▶ blaue, braune, graue ~n; →a. blau; funkelnde, glänzende, strahlende, sanfte, schöne, treue ~n; feuchte, nasse ~n bekommen, haben (fast) zu weinen beginnen, sehr gerührt sein; gute, schlechte ~n haben gut, schlecht sehen; da blieb kein ~ trocken alle waren sehr gerührt: (umg.) begeistert; hast du keine ~n (im Kopf)? hast du das nicht gesehen?, pass doch auf!; mit offenen ~n in sein Unglück rennen das U. sehen, aber nicht beachten; jmdm. schöne ~n machen mit ihm kokettieren, seine Aufmerksamkeit zu fesseln suchen; sechs ~n werfen (beim Würfeln); ein sicheres ~ für etwas haben das Wesentliche von etwas erfassen imdm. etwas an den ~n ablesen, absehen einen unausgesprochenen Wunsch erraten; das passt wie die Faust aufs ~ (umg.) das passt gar nicht zusammen; imdm. etwas aufs ~ drücken ihm eine lästige Arbeit aufzwingen; ein ~ auf etwas werfen etwas gern besitzen wollen; er hat ein ~ auf sie geworfen er interessiert

blau ⟨Adj.; -er, am -(e)s|ten⟩ 1 von blauer Farbe, von der Farbe des Himmels bei Sonnenschein;

Zus blaugrau, blaugrün, blauschwarz 2 ⟨umg.; übertr.⟩ betrunken ▶ ~e Augen haben; er hat ein ~es Auge ⟨übertr.⟩ (durch Schlag od. Sturz) blutunterlaufenes Auge; der Himmel ist ~; sie trägt ein ~es Kleid; das ~e Meer; der Blaue Planet die Erde ▶ jmdn. grün und ~ schlagen heftig schlagen ▶ er war schon am Nachmittag völlig ~ ⟨übertr.⟩ ▶ mit einem ~en Auge davonkommen ⟨Sprichw.⟩ mit dem Schrecken od. einer nur geringfügigen Verletzung davonkommen; ~ vor Kälte sein sehr blass u. stark fröstelnd vor Kälte; →a. Brief

blond ⟨Adj.; -er, am -es|ten⟩ 1 hell, gelblich (Haar)
2 hellhaarig (Person) ▶ eine Frau mit ~em Haar;
das Kind hat ~e Locken; ein (kühles) Blondes
ein Glas helles Bier; Zus blondhaarig, blondlockig; dunkelblond, hellblond, rotblond

~ jetzt anders als früher > blondes, dunkles, graues, helles, rotes, schwarzes, weißes ~ haben; dichtes, dünnes, fettiges, glänzendes, glattes, krauses, lockeres, lockiges, schönes, schütteres, seidiges, starkes, strähniges, welliges, widerspenstiges ~; falsches, künstliches ~ Perücke; das

Haar (das; -(e)s, -e) I (zählb.) sehr feines, fadenförmiges Gebilde, das aus der Haut von Tieren
u. Menschen wächst | ausfallen: die ~e fallen
ihm aus; er kann niemandem ein ~ krümmen
(übertr.) niemandem etwas Böses tun; wir wollen hier keine ~e spalten (übertr.) wir wollen
uns nicht über unwesentl. Kleinigkeiten streiten;
die ~e stehen einem zu Berge man ist entsetzt;
die ~e sträubten sich ihm vor Schreck (übertr.)
| blonde, graue ~e, (besser) blondes, graues ~;
darüber lasse ich mir keine grauen ~e wachsen
darüber mache ich mir keine Sorgen; kein gutes



QUASTHOFF KOLLOKATIONENWÖRTERBUCH (2011) STARK QUANTITATIV UND TOOL-BASIERT

Zahn der

V: NOM. abbrechen · ausfallen · blitzen · herausragen · schmerzen · strahlen • sich lockern • AKK. behandeln · bekommen · polieren · putzen · reparieren · richten • (Pl.) entblößen · fletschen · zeigen · zusammenbeißen • ausbeißen · ausreißen · ausschlagen · einschlagen · ruinieren · schädigen · verlieren • DAT. in den Zähnen pulen • mit den Zähnen klappern · knirschen · mahlen

A: echt · gesund · kräftig · natürlich · perfekt · schön » entzündet · faul · faulig · hohl · kaputt · kariös · krank · krumm · locker · lose · schadhaft · schief · schlecht · stumpf · vereitert · vorstehend · wackelnd » abgebrochen · angegriffen · ausgefallen · ausgeschlagen · herausgeschlagen » dritte · falsch · künstlich » blitzend · blütenweiß · braun · gelb · geputzt · golden · perlweiß · schneeweiß · schwarz · strahlend · weiß » knirschend · zusammengebissen · zusammengekniffen » gefährlich · messerscharf · riesig · scharf · spitz



BUHOFER FESTE WORTVERBINDUNGEN (2014) (PRINT)

Auge i

ADJEKTIVE/ADVERBIEN

trübe | dunkle | braune | blaue Augen Außer mir haben in meiner Familie alle blaue Augen. | grüne | gelbe Augen Die Katze schaute ihn aus großen, gelben Augen an. | helle | klare Augen ■ das rechte∘ | linke A.∘ Er hat eine Netzhautablösung am linken Auge erlitten. ■ mandelförmige∘ | große Augen ■ glänzende Augen∘ Das Mädchen schaute mit glänzenden Augen durch die Schaufensterscheibe des Spielzeugladens. | leuchtende∘ | funkelnde Augen Mit funkelnden Augen erzählen sie von ihren Plänen und Visionen. | glühende∘ | blitzende∘ | feurige∘ | stechende Augen∘ ›durchdringender Blicke ■ schöne | aufmerksame | kluge Augen∘ Der Vogel blinzelte sie mit seinen

schwollenes | blaues A. Bluterguss am Auges Gestern hat er sich mit einem Fremden geprügelt und hat jetzt ein blaues Auge. | blutunterlaufene Augen | schielendes A. Er

blau

ADJEKTIVE/ADVERBIEN

■ leuchtend b. Aufmerksam betrachtete sie ihn mit ihren leuchtend blauen Augen. | strahlend b. Ein strahlend blauer Himmel und 30 Grad im Schatten – so gefällt mir der Sommer. ■ herrlich | schön | wunderschön b.

Nomen

b. wie der Himmel

ZUSAMMENSETZUNGEN

- blassblau | hellblau | blitzblau A Den blitzblauen
 Sommertag hätte er besser nutzen können, als in diesem
 Konferenzsaal herumzusitzen. | knallblau | tiefblau |
 dunkelblau azurblau | himmelblau | indigoblau |
 kobaltblau | königsblau | lichtblau | marineblau |
 nachtblau | saphirblau | stahlblau Auf dem Bild sieht man
 drei Eichen unter stahlblauem Himmel. | taubenblau |
 türkisblau | veilchenblau | wasserblau graublau |
 grünblau | rotblau | schwarzblau
- → weitere Artikel mit blau:
- Auge | Brief | Farbe | Fläche | Fleck | Hemd | Himmel | Lampe |
 Licht | Lippe | Meer | Rauch | Schild | See | Uniform | Zelt | Zone
- blühen | markieren | schlagen grau | grün | rot | schwarz



BUHOFER FESTE WORTVERBINDUNGEN (2014) (PRINT)

Zahn m

ADJEKTIVE/ADVERBIEN

kräftige | saubere | gesunde Zähne Ihre gesunden Zähne sind das Ergebnis täglicher Pflege. | schöne | spitze | stumpfe | scharfe | messerscharfe Zähne Der Hai hatte messerscharfe Zähne, die sich bei dem Angriff in den Schiffsrumpf bohrten. | vorstehende | schiefe Zähne Schiefe Zähne sind nicht immer nur angeboren, sondern werden häufig durch Karies verursacht. weiße | gelbe Zähne^o Der Konsum von Nikotin verursacht gelbe Zähne. | schwarze Zähne^o Die Folgen von Karies sind schwarze Zähne. | schlechte Zähne | lockerer | kranker | schmerzender | kaputter | hohler | fauler Z. Wenn sie lächelt, sieht man die faulen Zähne. die ersten Zähne^o Mit sechs Jahren hat man noch die ersten Zähne. | die zweiten Zähne | bleibender Z.° | die eigenen Zähne° | natürliche | echte | künstliche | falsche Zähne Falsche Zähne haben den Vorteil, dass man sie zum Reinigen herausnehmen kann. | die dritten Zähne >künstliche Zähne

VERBEN

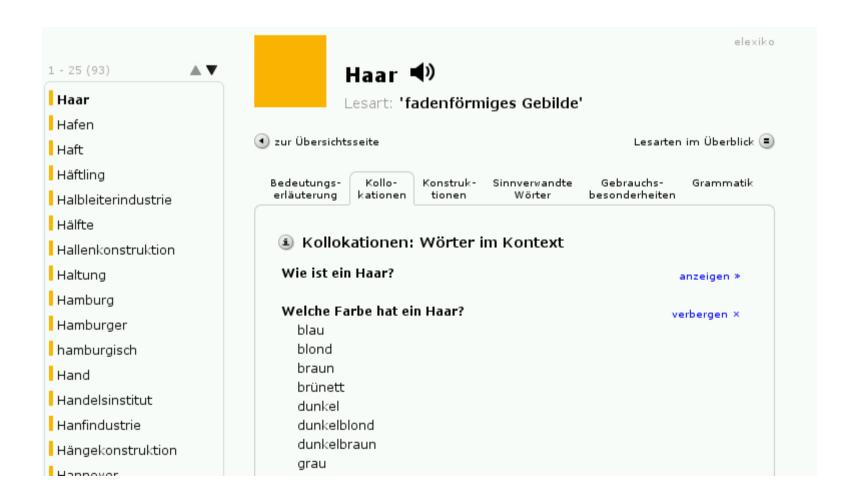
bekommen die ersten/zweiten Zähne **kommen** | ein Z. **bricht durch** Mit etwa 6 Jahren brechen bei den Kindern die ersten
bleibenden Zähne durch. | Zähne **bekommen**

verlieren

die Zähne. | die Zähne entblößen | mit den Zähnen knirschen Wer im Schlaf mit den Zähnen knirscht, schädigt auf Dauer das Gebiss. | durch die Zähne pfeifen | beißen mit den Zähnen nach etw. schnappen | etw. mit den Zähnen zerbeißen | mit den Zähnen an etw. nagen Aus Nervosität nagte er mit seinen Zähnen an der Unterlippe. | die Zähne zusammenbeißen° 1. Vor Anstrengung biss der Gewichtheber die Zähne zusammen. | 2. übertr. stapfer sein, durchhalten: | die Zähne in etw. graben° | schlagen° Gierig vor Hunger schlugen wir die Zähne in das frische Brot. Inflegen Zähne blitzen Aus dem strahlenden Gesicht blitzten weiße Zähne. (sich) die Zähne putzen° | die Zähne reinigen° Einmal jährlich werden die Zähne von der Dentalhygienikerin intensiv gereinigt. | in den Zähnen stochern Ich stocherte mit einer Nähnadel in den Zähnen, um die Fleischreste herauszubekommen. | sich die Zähne machen lassen In Westeuropa ist es Mode, sich in Osteuropa billig die Zähne machen zu lassen. ausreißen einen Z. ziehen ausziehen reißen Bei meinem Zahnarzt wurde der faule Zahn gerissen. | einen Z. ausreißen | entfernen = einschlagen imdm. die Zähne ausschlagen | einschlagen Im Streit schlug der eine Mann dem anderen die Zähne ein.



IDS "ELEXIKO" (NUR ONLINE)





elexiko

1 - 25 (93) **▲ ▼** Hafen Haft Häftling Halbleiterindustrie Hälfte Hallenkonstruktion Haltung Hamburg Hamburger hamburgisch Hand Handelsinstitut Hanfindustrie Hängekonstruktion Hannover hart häufia Haupt Hauptstadt Hausgeräteindustrie Haushalt Haushaltskontrollausschus Haushaltskontrolle Hausindustrie



Haar ◆)

Lesart: 'fadenförmiges Gebilde'

zur Übersichtsseite

Lesarten im Überblick

Bedeutungserläuterung

Kollokationen Konstruktionen

Sinnverwandte Gebrauchsbesonderheiten Wörter

Grammatik

Konstruktionen: Typische Verwendungen

Verwendungen mit Attribut

verbergen ×

[Eigenschaft: z. B. blond] gefärbtes Haar

nach hinten gekämmtes Haar

noch kein einziges graues Haar

Falten und graue Haare

rote Haare und Sommersprossen

[Eigenschaft: Z. B. dunkle] Haare

[Eigenschaft: z. B. graues, schütteres] Haar

[Eigenschaft: Z. B. kurzes], [Eigenschaft: Z. B. blondes] Haar

[Eigenschaft: Z. B. lange], [Eigenschaft: Z. B. braune] Haare

[Eigenschaft: Z. B. dunkles] Haar und [Eigenschaft: Z. B. braune]

Augen

[Eigenschaft: Z. B. normale] Figur und [Eigenschaft: Z. B. hellblonde]

▲ ▼

ein [Eigenschaft: z. B. schmales] Gesicht und [Eigenschaft: z. B. schwarzes] Haar

das Haar zu einem Pferdeschwanz gebunden

die Haare zu einem Zopf gebunden

die Haare nach hinten gekämmt

das Haar [Eigenschaft: z. B. ordentlich] gekämmt

das Haar [Eigenschaft: z. B. streng] gescheitelt

die Haare unter einem Kopftuch versteckt

mit blonden Haaren und blauen Augen



KOOKKURRENZANALYSE DES IDS

- Kookkurrenzanalyse des IDS (C. Belica, 1995)
 - seit 1995 online weltweit abrufbar
 - auf beliebige virtuelle Korpora anwendbar
 - derzeit in COSMAS II integriert
 - seit 2003 Weiterentwicklung
- einstellbare Kontextgröße
- Autofokus (positionelle Gebundenheit)
- Zuverlässigkeit, Granularität
- optional Lemmatisierung
- optional Synsemantika
- syntagmatische Muster
- ein- oder mehrdeutige Clusterzuordnung
- mehrere Statistiken (log-likelihood ratio LLR, mutual information MI ...)



BEISPIEL: WEICHSEL

- neben Fundus für illustrierende Angaben auch:
- (lexikografische) Lesartenoder Bedeutungsdisambiguierung anhand Kookkurrenzprofil
- eröffnen Vergleichsmöglichkeiten zwischen verschiedenen Wörtern

5 -5 -2 369 östlich unteren	4 100% östlich der unteren Weichsel
<u>▼</u> -5 -2 369 östlich gelegenen	4 100% östlich der Weichsel gelegenen
<u>■</u> -5 -2 369 östlich	33 100% östlich der Weichsel
3 3 335 Memel Land	7 71% Land zwischen Oder Neiße Weichsel und Memel
-3 3 335 Memel	21 80% Oder zwischen Weichsel [und] Memel
5 4 299 Kirschen Pfirsiche Birnen	3 66% Birnen [] Kirschen [] Weichseln Pfirsiche
5 4 299 Kirschen Pfirsiche	4 50% Kirschen [] Weichseln Pfirsiche
▶ -5 4 299 Kirschen Himbeeren	4 50% Himbeeren [] Weichseln Kirschen
■ -5 4 299 Kirschen	25 56% Kirschen [und] Weichseln
-3 -2 242 Flutwelle Pommern erreichte	1 100% Flutwelle Weichsel erreichte Pommern
-3 -2 242 Flutwelle Pommern	3 100% Die Flutwelle der Weichsel durch Pommern
-3 -2 242 Flutwelle erreichte	4 100% Die Flutwelle der Weichsel [] erreichte
-3 -2 242 Flutwelle	19 94% Die Flutwelle der Weichsel am
■ -4 -2 221 Land	49 91% das Land zwischen an der Weichsel und
■-5 4 194 Nogat	10 60% der Weichsel [in und Weichsel] Nogat
-5 3 190 Thorn Wasserstand	3 100% Wasserstand der Weichsel im nordpolnischen Thorn Torun
■ -5 3 190 Thorn geboren	4 100% 1940 in Thorn an der Weichsel [] geboren
■ -5 3 190 Thorn	13 61% 1940 in Thorn an der Weichsel geboren
-3 -3 190 Wunder	20 100% das Wunder an der Weichsel
3 3 188 Zuflüsse	12 91% Pegelstände der Weichsel und ihrer Zuflüsse sanken
-1 -1 186 Manfred Staatsrat	9 77% Staatsrat [] Manfred Weichsel
■ -1 -1 186 Manfred	22 100% Staatsrat Manfred [] Weichsel
3 4 171 Wloclawek geboren	4 100% 1920 in Wocławek an der Weichsel [] geboren
-3 4 171 Wloclawek	9 77% 1920 in Woclawek an der Weichsel geboren
-5 5 169 Warschau	22 77% an der Weichsel in von Warschau und
💶 -4 4 159 Staatsrat	15 80% Staatsrat [Manfred] Weichsel der
5 152 Pfirsiche Zwetschken Pflaumen	2 100% Weichseln Pfirsiche Zwetschken und Pflaumen
5 5 152 Pfirsiche	12 66% Weichseln [zu S kg] Pfirsiche und



VERGLEICH VON KOOKKURRENZPROFILEN KOOKKURRENZDATENBANK CCDB

Elbe, Neiße, Donau, Fluss, Fluß, Wolga, Neckar, Spree, Rhein, Euphrat

östlich (7) Weser (7) Rhein (9)
Hochwasser (7) Donau (8) südlich (7)

Spree (3) Nordsee (2) Nebenfluss (7) Ufer (10)
Wasserstand (6) nördlich (9) Moldau (2)
Pegel (8) fließt (10) mündet (9)
Wasser (10) Flüsse (8)
Mündung (8) Oberlauf (7) Pegelstand (4)

Nebenflüsse (4) Unterlauf (7) Stadt (10) Deiche (4)

Meter (4) Flüssen (7) Deich (1) gelegenen (7)
Pegelstände (6) Ostsee (3) Neiße (2) Flutwelle (2) Gebiete (3)

Brücke (10)

Kilometer (7) unteren (5)

Weichsel ⇔ ähnliche Profile �

Marille, Himbeere, Birne, Zwetschge, Kirsche, Johannisbeere, Pfirsich, Pflaume, Aprikose, Brombeere

Pfirsiche (9) Kirschen (9)

Zwetschken (7) Pfirsich (9) Birnen (9)

Äpfel (10) Nektarinen (9)

Erdbeeren (10)

Himbeeren (9) Pflaumen (9)

Kilo (9) Karfiol (1) Ribisel (5) entkernen (4)

Kirsche (9) 40 (5) Äpfeln (10)

35 (4) Birne (9) Zwetschke (7) Süßkirschen

(7) Sauerkirschen (9) Apfel (10)

Topfen (1) 50 (5) 100 (5) 70 (1)

45 (3) Waschen (8) Sorten (8) 38 (2) 60 (2) waren (2)

wurden (1) 48 (3) 12 (3) allem (3) Sind (10)



VERGLEICH VON KOOKKURRENZPROFILEN KOOKKURRENZDATENBANK CCDB © Cyril Belica: Modelling Semantic P

- hohe Ähnlichkeit im Kookkurrenzverhalten liefert eine Liste von Wörtern, die Kandidaten für verschiedene lexikalische Relationen sein können
 - Kohyponymie
 - Synonymie
 - Antonymie
 - Paronymie





SOM-KARTIERUNG ZU *WEICHSEL*

- "Self-organizing Map" der Liste der ähnlichen Profile
- Gruppierung und Anordnung gemäß Stärke der Ähnlichkeit untereinander
- verschiedene "Regionen"
 lassen sich als verschiedene
 Lesarten (hier: sogar
 Bedeutungen) deuten

© Cyril Belica: Modelling Semantic Proximity - Self-Organizing Map (version: 0.32, init tau: 0.04, dist: u, iter: 10000)

Weichsel

	Aprikose Apfel Erdbeere Weintraube Rhabarber dag Melone Feige		Oder Schneeschmelze Hochwasseralarm Hochwassermarke längs Hochwasserlage Jahrhunderthochwasser	Unterlauf Pegelstand Wasserstand Pegel Oberrhein Schifffahrt Schiffahrt Fließgeschwindigkeit	Elbe Rhein Neckar Mosel Saale Nebenfluß Nebenfluss Theiss
	Birne Himbeere Marille Pflaume Pfirsich Kirsche Zwetschge Johannisbeere	Steige Obstgarten	Hochwasser Hochwassergebiet Deich fluten Sturmflut Flutwelle Flut Regenfall	Oberlauf Jangtse Stausee Fahrrinne ertrinken Schiffsverkehr Staudamm Flussbett	Donau schiffbar Isar Weser Mississippi Rhone münden Salzach
		zerstört	Überschwemmungsgeb schwemmen überschwemmt überschwemmen überfluten überflutet schwappen spülen	Fluss Fluß flussabwärts flußabwärts Mulde Ufer Bache entlang	Spree Mekong Mündung Themse Zusammenfluß Zusammenfluss durchfließen Flüsschen
	Danzig Warschau Krakau Stettin Warschauer polnisch Schlesien Pommern	arschau Stettiner Moldau akau Tatra Lauenburg ettin Ostgrenze Oderbruck arschauer Inisch hlesien		Grenzfluß Grenzfluss beiderseits Ostsee sumpfig Flußufer Flussufer	Wolga Tigris Euphrat Dnjepr Wasserscheide Meer Loire Flußmündung
	Polen Galizien Pole Ostpreußen Pol Ostpreuße deportieren Kiew	Bukowina Westgrenze Sibirien Ukraine Kaliningrad Krim Bessarabien Kaukasus	Dnjestr Karpaten Ural Ostseeküste Haff Torgau umsiedeln Baikalsee	ostwärts östlich nahen nahe gelegen unweit nördlich südlich	Landzunge Bucht Westufer Tiefebene Atlantik Nahe Ostufer Hügellandschaft



CNS WEICHSEL VS. RHEIN

- "Contrast Near-Synonyms"
- Anwendung des gleichen Verfahrens auf Paare von Wörtern
- zusätzliche Einfärbung im Übergang von gelb zu rot entsprechend Nähe zu den vorgegebenen Wörtern
- orange indiziert geteilte Verwendung, Primärfarben diskriminierende Verwendungen

© Cyril Belica: Modelling Semantic Proximity - Contrasting Near-Synonyms (version: 0.21, init tau: 0.4, dist: x, iter: 10000)

Weichsel Rheir	า			
Birne Himbeere Marille Pflaume Pfirsich Aprikose Kirsche Zwetschge	Obstgarten	ostwärts östlich nahen nahe gelegen Bucht Landesinnere Ural	Karpaten Tatra Haff Westgrenze Ostgrenze Sibirien Riesengebirge Ostseeküste	Danzig Polen Galizien Warschau Pole Krakau Stettin Warschauer
Steige	überschwemmt Hochwassergebiet überschwemmen Flutwelle überfluten überflutet Regenfall schwappen	Lauenburg Westufer Landstrich Ostufer Hügellandschaft Oderbruch Hügel durchquert	Neiße Dnjestr Moldau Wolga Dnjepr Tigris Loire Grenzfluss	Stettiner
Saale Amazonas Wupper Werra Aare Maas entspringen Geografie Mosel Lahn Oberrhein Main Nahe Saar Ruhr Vogelsberg	Jangtse Hochwasser Mekong Theiss Mississippi Rhone Deich Ganges Oberlauf Neckar Unterlauf Pegelstand Mündung Wasserstand Nebenfluß Nebenfluss	Mulde durchfließen entlang Flüsschen Flusslauf Flußlauf Wasserscheide entwässern Holzbrücke Renaturierung Kanal renaturieren Dorfbach Altarm kanalisiert Flußbett	Elbe Donau Fluß Fluss Spree schiffbar flußabwärts flussabwärts Fahrrinne Schleuse ertrinken Schiffsverkehr Schifffahrt Wasserstraße schippern Binnenschiff	Themse Meer Sandbank Maggiore Bodensee Atlantik Ostsee Flussmündung Hafenbecken Mittelmeer Eisscholle Ponton Landwehrkanal flussaufwärts Meerenge flußaufwärts
Koblenz Mittelrhein Bingen Rüdesheim Rheinhessen Binge Loreley Niederrhein	Rheinbrücke rechtsrheinisch Rheinseite Rheinufer rheinabwärts Lindwurm Limes Tagesausflug	Wasserqualität Kühlwasser Rinnsal Uferbereich Fischart reißend Kloake Wassermenge	Gewässer Gestade Baggersee gesprungen schwimmen paddeln Südufer Teich	Zürichsee Untersee Wasserschutzpolizei See Ärmelkanal Anlegestelle Motorboot Vierwaldstättersee



RESUMEE

- geeignete Datengrundlage vorausgesetzt, bieten Korpora Hinweise über
 - Belegrecherchen
 - Bewertung der Auffälligkeit der Treffermengenverteilung (Metadaten!)
 - insbesondere statistische Bewertung der Kontexte (typische Verwendungsweisen)
 - Differenzierung der Verwendungskontexte (Lesarten)
 - Beziehungen zwischen Wörtern anhand Vergleich der Verwendungskontexte (z.B. Synonymie u.Ä.)
- immer vor dem Hintergrund
 - der Zusammensetzung der Daten
 - der technischen Aufbereitung (etwa durch Tagging je nach Theorie und Implementierung)
 - und der Fragestellung:
 Welche Information möchte ich für welchen Nutzer in welcher Nutzungssituation wie anbieten?

VIELEN DANK

KORPUSLINGUISTIK@IDS-MANNHEIM.DE WWW.IDS-MANNHEIM.DE

Mitglied der Leibniz-Gemeinschaft