

# Textkorpora als Ressourcen für die Digital Humanities: Chancen, Herausforderungen, Perspektiven

Ringvorlesung „Digital Humanities: Die digitale Transformation der Geisteswissenschaften“ (Berlin, WS 2017/18)  
19. Dezember 2017



**Prof. Dr. Angelika Storrer**

UNIVERSITÄT  
MANNHEIM

# Überblick

- Chancen: Wie lassen sich digitale Korpusressourcen heute schon nutzen? (Fokus auf dem Deutschen und den Ressourcen der CLARIN-D-Zentren BBAW und IDS)
- Herausforderungen: In welchen Bereichen gibt es Einschränkungen / Grenzen? Was ist bei der Interpretation automatisch erzeugter Korpusauswertungen zu beachten?
- Perspektiven: Wünsche und Anregungen für die weitere Entwicklung von Ressourcen und Werkzeugen.

Weihnachtsmann



<https://www.dwds.de>



## Wörterbücher

465 000 Einträge in verschiedenen Wörterbüchern.

### Konfekt, das

**Geschlecht:** Substantiv (Neutrum) – Genitiv Singular: *Konfekts, Konfektes* – Plural: *Konfekte* – wird nicht im Singular verwendet

**Asynthese:** *ne*

**Wortbildung:** Konfekt (veranschaulicht)

**Herbst:** Latein

**Wortbildung:** mit *Konfekt* als Entlehnung [Konfektbrotchen](#) – mit *Konfekt* als Entlehnung [Eisbonbon](#)

### Bedeutungen

#### 1. Pralinen o. Ä.

**Beispiel:**

Im Hotel «Köln» stand *Konfekt* aus Marzipan [...] auf dem Substantiv [...] (St. York, 01.07.2002)

Auf dem Tisch stehen zur Begrüßung eine Fruchtstube

[Bsp.: Konfekt \(Präparat\), 1891](#)



## Textkorpora

Dreizehn Milliarden Belege aus historischen und gegenwartssprachlichen Textkorpora.

1-5 von 1217 Texten

in Harnel ans Himmell sein wahres <b>Gold</b>	Foll' oder selts Gold Wölle/ auf Erd
sein wahres Gold redV oder selts <b>Gold</b>	welle/ auf Erden herab.
stund stoffs der Kohren schwarz <b>Gold</b>	.
Du streust <b>Gold</b>	von deinen Höfen/ unsre Wesen an
Wen solts/ der Rubin in diesem <b>Gold</b>	/ nicht loben?
der andre macht verliebt/ ist von <b>Gold</b>	/ und scharf gespizet.
entz; (bleib) das mehr als göldne <b>Gold</b>	ihrer Haare/ ledig fliegen.
brüsten? daß mich solts/ wie nach <b>Gold</b>	/ nach dir gelösten?
Dessen Haupt von feinem <b>Gold</b>	/ die Brust und Arm von Silber/ der

## Statistische Auswertungen

Häufigkeiten, typische Verbindungen und zeitliche Verläufe über 400 Jahre.

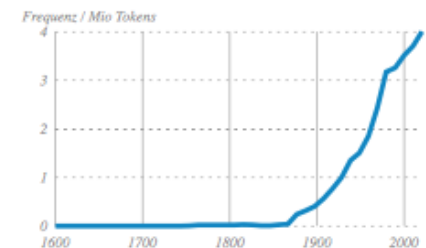


Weihnachtsmann

[Startseite](#)/ Wortinformation zu  
„Weihnachtsmann“

## Worthäufigkeit

selten häufig



## Ältere Wörterbücher

- [Grimmsches Wörterbuch \(1DWB\)](#) (1)
- [Wörterbuch der deutschen Gegenwartssprache \(WDG\)](#) (1)

## Korpustreffer

Referenzkorpora

- [DWDS-Kernkorpus \(1900–1999\)](#) (164)

# Weihnachtsmann, der

**Grammatik** Substantiv (Maskulinum)**Aussprache** **Worttrennung** Weih-nachts-mann (computergeneriert)**Wortzerlegung** ↗ [Weihnacht](#) ↗ [Mann](#)

Bedeutung

Thesaurus

Typische Verbindungen

Verwendungsbeispiele

eWDG, 1977

## Bedeutung

- zu Volksbräuchen um die Weihnachtszeit gehörige männliche Gestalt in pelzverbrämter Kleidung mit Rute und einem mit Geschenken gefüllten Sack

BEISPIELE:

hat dir der **Weihnachtsmann** viel gebracht?das Kind glaubt noch an den **Weihnachtsmann****Weihnachtsmänner** aus Schokoladeich sagte, daß der **Weihnachtsmann** es erfahre, wenn er seine Milch nicht trinke [NOSSACK, *November*, 147]

BEISPIEL:

Du bist genauso ein **Weihnachtsmann** wie die anderen [APITZ, *Nackt*, 312]

# Übersicht über die Korpora im DWDS

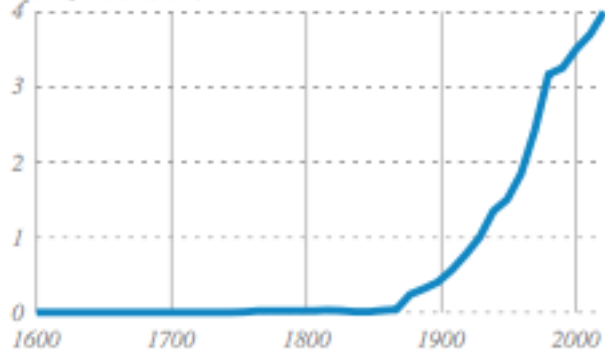


Korpus	Kategorie	Recherchierbarkeit	Dokumente	Tokens
DWDS-Kernkorpus (1900–1999)	Referenzkorpora	frei recherchierbar	79 116	121 397 604
DWDS-Kernkorpus 21 (2000–2010)	Referenzkorpora	frei recherchierbar	12 186	15 462 297
Deutsches Textarchiv (1473–1927)	Referenzkorpora	frei recherchierbar	4 014	210 748 369
Berliner Zeitung (1946–1993)	Zeitungskorpora	nach Anmeldung	1 570 943	497 816 459
Berliner Zeitung (1994–2005)	Zeitungskorpora	frei recherchierbar	851 194	237 059 321
neues deutschland (1946–1990)	Zeitungskorpora	nach Anmeldung	1 457 540	443 980 337
Tagesspiegel (1996–2005)	Zeitungskorpora	frei recherchierbar	330 790	156 317 581
Die ZEIT (1946–2016)	Zeitungskorpora	frei recherchierbar	1 202 804	548 347 820
Blogs	Spezialkorpora	frei recherchierbar	231 150	102 515 727
Webkorpus	Spezialkorpora	nach Anmeldung	6 978 183	3 012 009 215
Dortmunder Chat-Korpus	Spezialkorpora	nach Anmeldung	470	1 003 458
Filmuntertitel	Spezialkorpora	frei recherchierbar	12 008	75 599 650
Polytechnisches Journal	Spezialkorpora	frei recherchierbar	42 394	77 566 051
DDR	Spezialkorpora	frei recherchierbar	1 103	8 502 921
Gesprochene Sprache	Spezialkorpora	frei recherchierbar	600	2 858 964
<b>gesamt</b>			<b>12 774 495</b>	<b>5 511 185 774</b>

## Worthäufigkeit

selten  häufig

Frequenz / Mio Tokens



### Referenzkorpora

- | [DWDS-Kernkorpus \(1900–1999\)](#) (164)
- | [DWDS-Kernkorpus 21 \(2000–2010\)](#) (44)
- | [Deutsches Textarchiv \(1473–1927\)](#) (13)

### Zeitungskorpora

- | [Berliner Zeitung \(1994–2005\)](#) (2224)
- | [Tagesspiegel \(1996–2005\)](#) (915)
- | [Die ZEIT \(1946–2016\)](#) (1458)

### Spezialkorpora

- | [Referenz- und Zeitungskorpora \(aggregiert\)](#) (1679)
- | [Blogs](#) (682)
- | [Polytechnisches Journal](#) (0)
- | [Filmuntertitel](#) (724)
- | [Gesprochene Sprache](#) (1)
- | [DDR](#) (24)

Korpus:

DDR

Start:

1949

Ende:

1990

Anzeige:

KWIC  voll  maximal

Sortierung:

Datum absteigend

Anzahl Treffer pro Seite:

50

1–20 von 20 Treffern (24 insgesamt)

Treffer exportieren

1:	1989	kunstlexikon2	...geformt, ausgestochen oder gemodelt, z.B. Osterhasen,	<b>Weihnachtsmänner</b>	, Tiergestalten, ältere und neuzeitl.
2:	1988	scherzer_den	...s, aber lange diskutiertes Problem: die Hohlkörper, diese	<b>Weihnachtsmänner</b>	aus Schokolade, die es in keinem La
3:	1988	scherzer_den	...bt es wieder keinen Fotografen, der die Kinder mit einem	<b>Weihnachtsmann</b>	fotografiert.
4:	1987	kunstlexikon2	...läser ließ bald viele neue Gebilde entstehen, bunte Vögel,	<b>Weihnachtsmänner</b>	, Trauben, Nüsse, Knusperhäuschen
5:	1983	strittmatter_l	Mir brachte der	<b>Weihnachtsmann</b>	einen kleinen Spazierstock, seine Z
6:	1983	strittmatter_l	Es war vom	<b>Weihnachtsmann</b>	, schluchze ich.
7:	1983	strittmatter_l	...zu erkennen, daß die Laubsäge weniger ein Geschenk des	<b>Weihnachtsmannes</b>	als eines meiner Mutter war.
8:	1983	strittmatter_l	Der	<b>Weihnachtsmann</b>	hatte in jenem Jahr, von dem ich re
9:	1971	braun_helfen	Sehr beliebt in der Weihnachtszeit sind	<b>Weihnachtsmänner</b>	, die du vielleicht schon einmal geba
10:	1971	braun_helfen	... Gesicht, klebst einen großen Wattebart an und setzt dem	<b>Weihnachtsmann</b>	eine Zipfelmütze aus Buntpapier au
11:	1971	braun_helfen	Wie	<b>Weihnachtsmänner</b>	, Zitronenschweinchen, Hühnchen u
12:	1970	johnson_jahr	...e zum Aufbruch gedrängt, als das neue Vereinsmitglied	<b>Weihnachtsmann</b>	mit seinem Sack auftrat, so daß die
13:	1970	johnson_jahr	...ich denkt er, er kriegt die Präsidentschaftskandidatur vom	<b>Weihnachtsmann</b>	.
14:	1970	Briefe_Reima	...herum mit der beleidigten Miene von Kindern, denen der	<b>Weihnachtsmann</b>	keinen Komfort geschenkt hat, und

# Typische Verbindungen

computergeneriert



DWDS-Wortprofil



Detailliertere Informationen bietet das [DWDS-Wortprofil zu »Weihnachtsmann«](#).



- 1: Die Zeit, 03.12.2012, Nr. 49  
Der **Weihnachtsmann**, der **Osterhase**, Sandmann und die Zahnfee:
- 2: Die Zeit, 26.11.2012 (online)  
**Weihnachtsmann**, **Osterhase**, Zahnfee, Sandmann und der Mann im Mond müssen den Schwarzen Mann, Pitch, besiegen, um den Kindern das Staunen und die Fantasie zu erhalten.
- 3: Die Zeit, 28.03.2011, Nr. 13  
"**Weihnachtsmann** im Sinne des Gesetzes ist auch der **Osterhase**".
- 4: Die Zeit, 14.12.2009, Nr. 50  
Sein Vater machte noch **Weihnachtsmänner** und **Osterhasen**, sonst nichts.
- 5: Die Zeit, 17.12.2007, Nr. 51  
Der **Weihnachtsmann** ist kein **Osterhase**.
- 6: Die Zeit, 16.04.2007, Nr. 16  
Da wird doch eher der **Osterhase** zum **Weihnachtsmann**. "
- 7: Die Zeit, 29.01.2007, Nr. 05  
Der **Weihnachtsmann** ist eben doch ein **Osterhase**.
- 8: Die Welt, 24.12.2005  
Ein **Weihnachtsmann** mit einem **Osterhasen** und einem Glücksschwein, auf dem Rücken trägt er Silvesterraketen.
- 9: Die Welt, 20.10.2005  
Fast die Hälfte der Menschen in Deutschland will Kindern den Glauben an **Weihnachtsmann** und **Osterhasen** so lange wie möglich lassen.

# Korpora am Institut für Deutsche Sprache IDS in Mannheim

## DAS DEUTSCHE REFERENZKORPUS – DeReKo



32,83 Milliarden Wörtern (Stand 01.10.2017)

<http://www.ids-mannheim.de/cosmas2/>

## ***COSMAS II***

### Corpus Search, Management and Analysis System



#### Download-Server

Außerdem stehende folgende Korpora jeweils unter der CC-BY-SA-Lizenz zum Download zu Verfügung

- Reden und Interviews-Korpus (rei)
- Wikipedia-Korpora  
Aufbereitung 2011 in Kooperation mit dem Projekt EuroGr@mm [1],  
Aufbereitung 2013 und 2015 in Kooperation mit dem Programmbereich Forschungsinfrastrukturen [2].  
Aufbereitung 2017 durch den Programmbereich Korpuslinguistik.

#### DEUTSCHSPRACHIGE WIKIPEDIA - VERFÜGBARE FILES 2011-2017 (ENCODING ISO-8859-1)

Jahr	WP-Teilkorpus	I5	WikiXML	TreeTagger Standoff
2011	Artikel	wpd11.xces.bz2	-/-	-/-
	Artikeldiskussionen	wdd11.xces.bz2		
2013	Artikel	wpd13.i5.xml.bz2	dewikixml-20130728-articles.tar.gz	wpd13.tt.xml.bz2
	Artikeldiskussionen	wdd13.i5.xml.bz2	dewikixml-20130728-discussions.tar.gz	wdd13.tt.xml.bz2

# Datenbank für gesprochenes Deutsch (IDS Mannheim)



Korpuskennung	Korpusname	Erhebungszeitraum	Beschreibung
AD	Australiendeutsch	1966-1973	Interviews, Erzählungen und Bildbeschreibungen von und mit vorwiegend älteren Frauen und Männern, deren Familien z.T. schon seit drei Generationen in Südaustralien leben
BB	Deutsche Mundarten: Kreis Böblingen	1965-1970	Erzählungen und Unterhaltungen von und mit Sprechern unterschiedlichen Alters aus dem Kreis Böblingen
BR	Biographische und Reiseerzählungen	1985-1990	In der DDR aufgezeichnete Erzählungen und Interviews von und mit jungen Männern und Frauen
BW	Berliner Wendekorpus	1992-1996	Interviews mit 30 Ostberlinern und 26 Westberlinern (Frauen und Männer) im Alter von 19 bis 55 Jahren über die Themen Mauerfall 1989 sowie individuelle, soziale und wirtschaftliche Aspekte des Alltagslebens im Zeitraum 1992 bis 1995
DR	Deutsche Mundarten: DDR	1960-1968	Erzählungen, Unterhaltungen und Standardtexte (Vergleichstexte, Wortlisten), aufgenommen mit einheimischen Informantinnen und ÜbersiedlerInnen aus den ehemaligen deutschen Ostgebieten, die in Berlin, Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt und Thüringen wohnhaft waren
DS	Dialogstrukturen	1960-1977	Sprecher der Standardsprache bzw. standardnahen Sprache in Sprechereignissen unterschiedlicher Art
EK	Elizitierte Konfliktgespräche	1988-1990	Elizitierte Konfliktgespräche zwischen Müttern und jugendlichen Töchtern
FOLK	Forschungs- und Lehrkorpus Gesprochenes Deutsch	2003-2016	Alltagsgespräch, Institutionelle Kommunikation, Experimentsituation bzw. Kommunikationsspiel, Medien- bzw. öffentliche Kommunikation
FR	Grundstrukturen: Freiburger Korpus	1960-1974	Sprecher der Standardsprache bzw. standardnahen Sprache in Sprechereignissen unterschiedlicher Art
GWSS	Gesprochene Wissenschaftssprache	2009-2012	Universitäre Prüfungsgespräche, Vorträge von Experten und Studenten sowie Diskussionen dieser Vorträge



<https://dgd.ids-mannheim.de>

Teilkorpus	Titel	Anzahl Transkripte	~ Anzahl Types	~ Anzahl Tokens
FOLK	Forschungs- und Lehrkorpus gesprochenes Deutsch	534	72173	1952159

0368	FR	boah mir tut alles weh evelyn (.) °h	
0369		(0.33)	
0370	EG	wie vom sport jetzt ja [aber ++++++ hm]	
0371	FR	[ja ey ► ich hab] so krassen muskelkater °h un ich hab auch nie gedacht un ich bin echt so dann °h ich liege da so un denk mir nein es geht einfach nich mehr es geht nich mehr	
0372		(0.46)	
0373	FR	°h und °h die (.) frau die des training macht also des video die lacht auch immer so völlig hysterisch verrückt °h un dann sagt [sie immer so] sachen wie °h ja °h ja ihr denkt ich bin verrückt aber das hab ich schon oft gehört und s[olche] sachen	
0374	EG	[°hh ]	
0375	EG	[hm]	
0376		(0.34)	
0377	FR	°h ja ja (.) °h [aber es is so witzig]	
0378	EG	[ich find diese leu]te immer_n bisschen nervig (.)[°h ]	
0379	FR	[auf jeden] fall °h aber es erinnert mich an ähm (.) ((schmatzt)) °h	
0380		(0.36)	
0381	FR	an äh st starsip (.) starship troopers irgendwie sieht [sie so aus al]s wär sie so eine °h [von äh] kämpferin von den starship troopers un dann denk ich mir immer °h ja das is jetzt hier das ((Lachansatz))	
0382	EG	(((Lachansatz)))	Alltagsgespräch: Telefongespräch
0383	EG	(((Lachansatz)))	Zwei Freundinnen sprechen am Telefon über den vorausgegangenen Unfall des Katers, den Arztbesuch und den Unfall des Freundes von EG, ein Trainingsvideo auf YouTube, Zahnreinigung, Haarkreide, verschiedene TV-Serien, Wochenendpläne, Festivals, die Arbeit, eine Wetterwarnung, veganes Katzenfutter, Ohrenerkrankung, Musik, verschiedene Bands, Werbekampagne einer Zigarettenfirma, die Olympischen Spiele und die politische Bedeutung, eine Dokumentation über Doping, Freundinnen und deren Freundeskreis, die Planung des Abends und versuchen, einen Termin für ein Treffen zu finden.
0384		(0.42)	
0385	FR	°h ausbildungstraining [für die spacegemeinde]	
0386	EG	[trainingscamp ja]	



## Ereignis FOLK\_E\_00084

# Datenbank für gesprochenes Deutsch (IDS Mannheim)

<input checked="" type="checkbox"/>	4		FOLK_00048	LP		oh mein <b>gott</b> anita nein	Weiblich
<input checked="" type="checkbox"/>	5		FOLK_00048	LP		oh mein <b>gott</b>	Weiblich
<div style="border: 1px solid black; padding: 5px;"> <p>0213 LS unser (.) in der</p> <p>0214 (0.61)</p> <p>0215 AM was habt_n ihr für_n absturz ge[habt]</p> <p>0216 LP [oh] mein <b>gott</b></p> <p>0217 LS ((lacht))</p> <p>0218 AM wo denn</p> <p>0219 LP *hh (.) wir ham_en absturzabend gehab[t]</p> </div>							
<input checked="" type="checkbox"/>	6		FOLK_00048	LP		oh mein <b>gott</b>	Weiblich
<input checked="" type="checkbox"/>	7		FOLK_00048	LP		oh mein <b>gott</b>	Weiblich
<input checked="" type="checkbox"/>	8		FOLK_00050	LP		oh mein <b>gott</b> nein	Weiblich

## Metadaten

### Geschlecht

Wert	#Tokens: Treffer	#Tokens: Gesamt	Treffer rel.
Weiblich	21	729,543	0.0028785144%
Männlich	4	847,644	0.0004718962%
Nicht dokumentiert	1	4,297	0.0232720503%

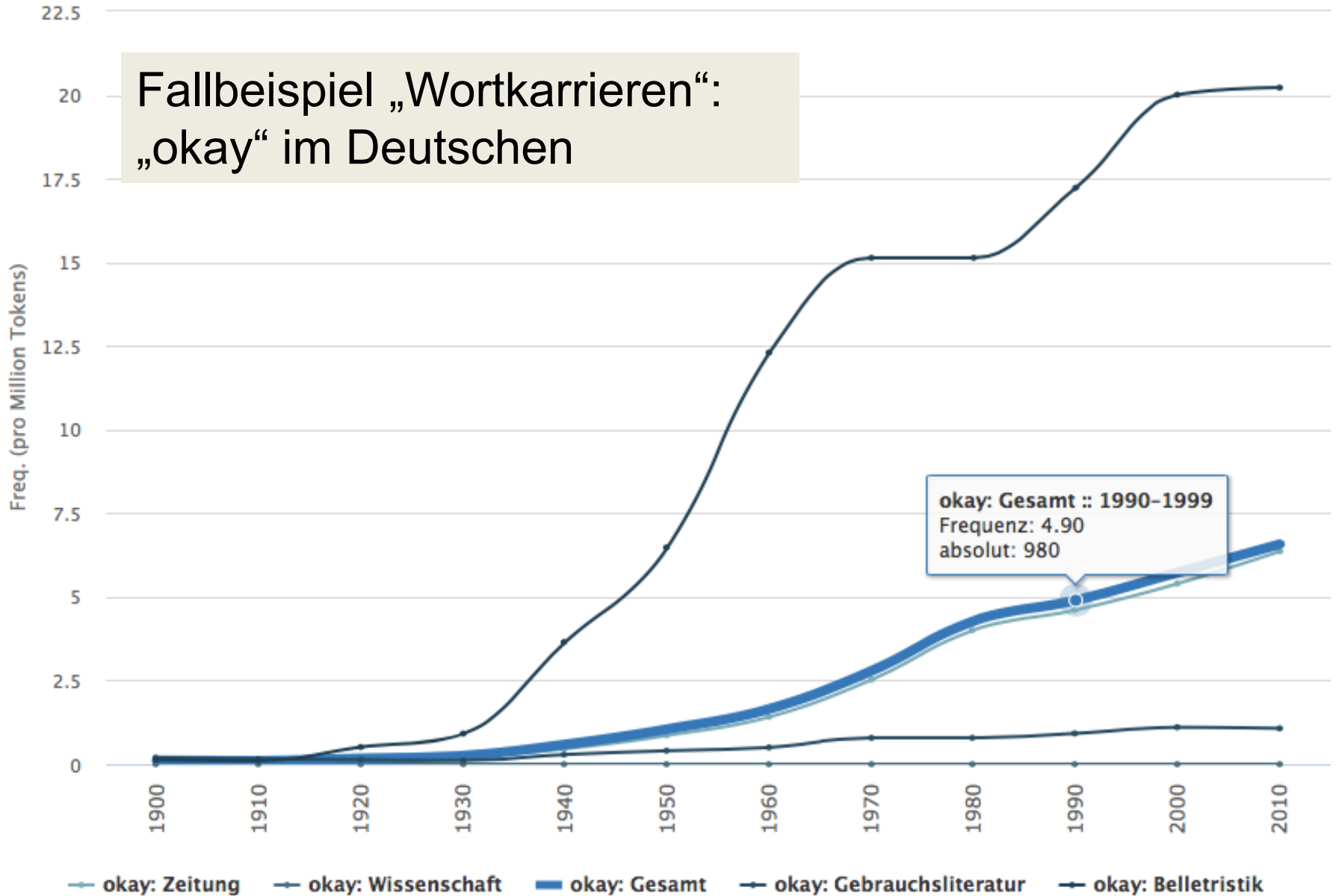
## Sprecher

Kennung	#Tokens: Treffer	#Tokens: Gesamt	Treffer rel.
FOLK_S_00184	8	13,503	0.0592460935%
FOLK_S_00041	1	5,773	0.0173220163%
FOLK_S_00182	1	40,625	0.0024615385%
FOLK_S_00343	1	2,689	0.0371885459%
FOLK_S_00178	1	16,243	0.0061564982%
FOLK_S_00388	1	8,046	0.0124285359%
FOLK_S_00403	1	755	0.1324503311%
FOLK_S_00401	1	3,815	0.0262123198%
FOLK_S_00363	1	36,087	0.0027710810%
FOLK_S_00301	1	4,773	0.0209511837%
FOLK_S_00424	1	4,485	0.0222965440%

<http://dgd.ids-mannheim.de/DGD2Web/ExternalAccessServlet?command=displayTranscript&id=FOLK E 00048 SE 01 T 02 DF 01&cID=c216&wID=w822&textSize=200&contextSize=4>

okay - Verlaufskurve

Fallbeispiel „Wortkarrieren“:  
„okay“ im Deutschen



»Man, Dadwolf, ich bin doch keine Rheuma-Oma...  
Ja, ich bin noch dran, Rike, Dad spielt mal wieder den  
Besorgten ... **okay**, ich komme nachher bei dir vorbei ...  
dann bis später ... Kußkuß!«  
Laura legt auf und wirft einen skeptischen Blick auf ihren  
Vater, der sich im Spiegel neben der Garderobe betrachtet.

Dückers, Tanja: Spielzone, Berlin: Aufbau-Taschenbuch-Verl. 2002  
[1999], S. 188

Nicht jeder gelistete Treffer ist ein Beleg für die Verwendung von „okay“ als deutsches Wort.

Einen Shanghaier Kuli, der vielleicht nur über einige Dutzend englische Wörter verfügt, **okay** sagen zu hören, ist nun ebenso alltäglich, wie z.B. einen armen russischen Handwerker in Shanghai, der nur für seine Landsleute arbeitet und nur notdürftig etwas Englisch radebricht, in rein russischer Konversation **okay** sagen zu hören.

Storfer, A. J.: PidGIN - ENGLISH. In: Gelbe Post, 01.05.1939, S. 18



# Chancen und Grenzen automatischer Annotationen

Linguistische Annotationen, z.B. die Zuordnung von Formen zu Wortarten (POS-Tagging), sind oft hilfreich, um die Präzision von Suchanfragen bei Homographen zu erhöhen.

**Beispiel:** Die Suche nach der Form „schicke“ als attributiv gebrauchtes Adjektiv (schließt Formen des Verbs „schicken“ aus).

In großen Korpora sind die linguistischen Annotationen, automatisch erzeugt. Die Qualität wird immer besser; bislang sind die automatisch annotierten Daten aber keineswegs fehlerfrei.

# Chancen und Grenzen automatischer Annotationen

## Suche nach der Wortform @schicke ohne POS-Tag



Ress

@schicke



DWDS Standardsicht

+Ressourcen



Kernkorpus 20



Treffer: 566, davon anzeigbar: 470

Im Kernkorpus wird zwischen Groß- und Kleinschreibung unterschieden.

- 1 1999 BE ...klingt es vom Zimmer auf den Gang hinaus. Ich **schicke** mich an, ins Zimmer zu gehen. Ich will nicht, daß sie bö...
- 2 1999 BE ...Du sollst das Inferno auf Erden erleben. Ich **schicke** dich in die Ofenhölle. Ofenhölle?...
- 3 1999 ZE ...Leser versonnen an. Im Innern präsentieren junge Designer **schicke** Bahnwagen der Zukunft, die irgendwie ganz anders aussehen al...
- 4 1998 ZE ... - wie der angesehene schwarze Essayist Stanley Crouch -, er **schicke** seine Musik auf den Strich, genauso wie er es zu Junkie-Zeit...
- 5 1998 ZE ...Für ihre Echtheit haben Sie mein Wort. Die Originale **schicke** ich heute an den SPD-Fraktionsvorsitzenden Rudolf Scharping,...
- 6 1997 ZE ...sen" - und wieder "Akquisen" an Werbekunden schreiben - "500 **schicke** ich pro Jahr mindestens raus" Und Michel? Willi Weber er
- 7 1997 ZE ... Die PR-Ladies, auch PR-Hasen genannt, schmeißen nicht nur **schicke** Partys, was allein schon "Knochenarbeit", so Alexandra von R...
- 8 1997 ZE ...üßes essen oder lieber später etwas Richtiges? Sofort die **schicke** neue Jacke kaufen oder doch für die Rente sparen? Heute P...
- 9 1996 BE ...erten Willi an die alten Frauen im Konsum, wenn sie über die **schicke** Pionierleiterin und ihre Männerbekanntschaften herzogen. ...
- 10 1996 BE ...Ich habe es leider eilig. Ich **schicke** dir ein Fernschreiben, codiert auf sieben. Halt mich auf ...
- 11 1996 ZE ... mehr losgelassen. Den lege ich hiermit endgültig ab, ich **schicke** ihn in die Wüste. Lange Perioden blieb er fern, tauchte g...
- 12 1994 BE ...nn bleibt die ja erst recht", regt sich Dionysos auf. Ich **schicke** ihr einen griechischen Mann, der sie schwängert", schlägt Ap...
- 13 1994 WI ... schlechte Schreibweise? paßt zu den Schuhen, die ich dir **schicke** «. Wie in der Redensart 'einen Stiefel vertragen' ist Stiefe...
- 14 1993 WI ... behaupten, der Stil der alten Maler des XVten Jahrhunderts **schicke** sich nicht mehr für die Kunst unserer Zeiten;... sein Beispi...
- 15 1990 GE ...le, 14. Mai 1990: » Sehr geehrter Herr Bundesminister, anbei **schicke** ich Ihnen eine Protokollnotiz über die Sie betreffenden Frag...
- 16 1988 BE ...ühlampenwerk für den Transport der Glühbirnen kein Auto hat, **schicke** ich meinen Fahrer... Gib mir das Ratsmitglied für Örtliche V...

# Chancen und Grenzen automatischer Annotationen

## Suche nach der Wortform @schicke als Adjektiv



Ress

\$p=ADJA with @schicke



DWDS Standardsicht

+Ressourcen



Kernkorpus 20



Treffer: 165, davon anzeigbar: 122

Im Kernkorpus wird zwischen Groß- und Kleinschreibung unterschieden.

- 1 1999 ZE ...Leser versonnen an. Im Innern präsentieren junge Designer **schicke** Bahnhöfe der Zukunft, die irgendwie ganz anders aussehen al...
- 2 1997 ZE ... Die PR-Ladies, auch PR-Hasen genannt, schmeißen nicht nur **schicke** Partys, was allein schon "Knochenarbeit", so Alexandra von R...
- 3 1997 ZE ...üßes essen oder lieber später etwas Richtiges? Sofort die **schicke** neue Jacke kaufen oder doch für die Rente sparen? Heute P...
- 4 1996 BE ...erten Willi an die alten Frauen im Konsum, wenn sie über die **schicke** Pionierleiterin und ihre Männerbekanntschaften herzo...
- 5 1994 WI ... schlechte Schreibweise? paßt zu den Schuhen, die ich dir **schicke** «. Wie in der Redensart 'einen Stiefel vertragen' ist Stiefe...
- 6 1986 BE ... die Haute Couture arbeitete. Sie nähte mir für ihn kleine **schicke** Anzüge. Die Hose hatte hinten stets ein Loch für den lang...
- 7 1986 GE ...aft abscheulichen) heimischen Klamotten angeödet und bekommt **schicke** Sachen im modernsten Styling. Datschenglück und Fernweh U...
- 8 1986 GF ...er Satz vor Freizeitartikeln, made in GDR, Taucheranzüge, **schicke** Bademoden, ein Surfbrett, Schlauchboote, Heimtrainer und wi...
- 9 1983 BE ... dir, was de willst! Und ich gehe weinend nach hinten und **schicke** meine Mutter nach vorn, damit sie notiert. Neuerdings gib...
- 10 1983 WI ...axen, Fakultäten, Kanzleien und Redaktionen. Eine gewisse **schicke** Bitterkeit untermalt sein Handeln. Denn Zyniker sind nich...
- 11 1971 GE ...hristlicher Gelassenheit erträgt; sie klauen ihm sogar seine **schicke** schwarze Fellmütze. Paul fällt mir dadurch auf, daß er de...
- 12 1971 GE ...ie schon Wünsche sind, sie da reproduzieren: Chef sein, eine **schicke** Sekretärin, mit der ein Verhältnis haben usw. Da punktuell a...
- 13 1971 GE ... Du freust dich über neue Sachen, über ein hübsches Hemd, **schicke** Schuhe, über ein besonders schönes Kleid. Ach, wie bald s...
- 14 1970 BE ... falle über Stühle, quetsche Finger, ich springe, ich hopse, **schicke** Telegramm » Hurra, bin angekommen « an Tanzstundenfreund auf...
- 15 1969 GF ...tzen (vol. das viel zu große Auto, die wenig praktische aber **schicke** Kleidung, den mäßig schmeckenden aber imponierenden Sekt). M...
- 16 1966 BE ...brigitte was mit dem elektriker anfängt den ich ihr manchmal **schicke** junger strammer kerl blauäugig gute germanische rasse nicht ...
- 17 1965 ZE ... Seiten fest, daß Großbritannien Verstärkungen nach Malaysia **schicke**, wie die USA immer neue Verstärkungen nach Südvietnam würfe...

# Bedeutung

▼ **umgangssprachlich**

1. **drückt Zustimmung aus** jawohl; es ist recht

BEISPIELE:

»wollen wir gehen?« »**Okay!**«

»konntest du alles erledigen?« »**Okay!**«

Interaktive Einheit

2. **〈etw. ist okay〉** etw. ist gut, ist in Ordnung

BEISPIELE:

wenn ich erst das Examen habe, ist alles **okay**

bei den Kindern war alles **okay**

adverbial bzw.  
prädikativ

- ▼ **〈jmd. ist okay〉** jmd. ist tüchtig, gesund

BEISPIELE:

der neue Mitschüler ist **okay**

er war zwei Tage krank, aber jetzt ist er wieder **okay**, fühlt er sich wieder **okay**

@okay with \$p=ADJD|

D

W

D

S

Korpus:

Referenz- und Zeitungskr

Start:

1900

Ende:

2016

Textklassen:

- Belletristik  Wissenschaft  Gebrauchsliteratur  
 Zeitung

Anzeige:

- KWIC  voll  maximal

Sortierung:

zufällig

Anzahl Treffer pro Seite:

50

1–50 von 118 Treffern (173 insgesamt)

Treffer exportieren

← -10 -5 ← 1 2 3 → +5 +10 →

- 1: Goosen, Frank: Liegen lernen, Frankfurt am Main: Eichborn AG 2000, S. 111  
Sie sagte **okay**, und wir gingen hoch.
- 2: Regener, Sven: Herr Lehmann, Frankfurt am Main: Eichborn AG 2006 [2001], S. 46  
»Das sind Spiegeleier mit Speck und Bratkartoffeln, das ist doch **okay**.«
- 3: Krausser, Helmut: Eros, Köln: DuMont 2006, S. 308  
Ist das **okay** für dich?«
- 4: Dückers, Jürgen: Spielzone, Berlin: Aufbau Taschenbuch Verh. 2002 [1999], S. 179  
»Okay, **okay**«, murmelt Laura.
- 5: Goosen, Frank: Liegen lernen, Frankfurt am Main: Eichborn AG 2000, S. 36  
Auch Dylan war **okay**, schon wegen »With God on our side«, wo er gegen die Deutschen wettete.
- 6: Menary, Sverker: Der Tod des Märchenprinzen, Hamburg: Bantam Verh. 1986 [1986], S. 281  
Wenn ich aufgetakelt mit knallengen Hosen und tief ausgeschnittenem Dekollete rumlaufen würde, dann brauchte ich mich nicht zu wundern, **okay**.

PTKANT    Antwortpartikel    ja, nein, danke, bitte

# Korpusbelege (Referenz- und Zeitungskorpora (aggregiert))

@okay with \$p=PTKANT 🔍 ?

**Korpus:**

Referenz- und Zeitungskr ▾

**Start:**

1900 ▾

**Ende:**

2016 ▾

**Textklassen:**

- Belletristik
- Wissenschaft
- Gebrauchsliteratur
- Zeitung

**Anzeige:**

- KWIC
- voll
- maximal

**Sortierung:**

zufällig ▾

**Anzahl Treffer pro Seite:**

50 ▾

Es wurden keine Treffer in diesem Korpus gefunden.

ITJ      Interjektion      mhm, ach, tja

# Korpusbelege (Referenz- und Zeitungskorpora (aggregiert))

@okay with \$p=ITJ



**Korpus:**

Referenz- und Zeitungskr ▾

**Start:**

1900 ▾

**Ende:**

2016 ▾

**Textklassen:**

- Belletristik
- Wissenschaft
- Gebrauchsliteratur
- Zeitung

**Anzeige:**

- KWIC
- voll
- maximal

**Sortierung:**

zufällig ▾

**Anzahl Treffer pro Seite:**

50 ▾

Es wurden keine Treffer in diesem Korpus gefunden.

@okay



## Korpus:

Referenz- und Zeitungskt

## Start:

1900

## Ende:

2016

## Textklassen:

- Belletristik  Wissenschaft  Gebrauchsliteratur  
 Zeitung

## Anzeige:

- KWIC  voll  maximal

## Sortierung:

zufällig

## Anzahl Treffer pro Seite:

50

1-50 von 118 Treffern (173 insgesamt)

Treffer exportieren

←	-10	-5	←	1	2	3	→	+5	+10	→
---	-----	----	---	---	---	---	---	----	-----	---

- 1: Goosen, Frank: Liegen lernen, Frankfurt am Main: Eichborn AG 2000, S. 130  
Ich fragte, was das Theater mache, und sie sagte, das sei auch **okay**.
- 2: Schulze, Ingo: Simple Storys, Berlin: Berlin-Verl. 1998, S. 259  
»Sag ihnen morgen früh, daß alles **okay** ist, ja?
- 3: Goosen, Frank: Liegen lernen, Frankfurt am Main: Eichborn AG 2000, S. 111  
Sie sagte **okay**, und wir gingen hoch.
- 4: Merian, Svende: Der Tod des Märchenprinzen, Hamburg: Buntbuch-Verl. 1980 [1980], S. 111  
Ich habe nach dem Dienstag abend, wo du mir das so beiläufig gesagt hast und nachher sehr hartnäckig darauf bestanden hast, daß das ganz unwesentlich für eine Beziehung sei, erstmal gedacht, **okay**, er hat Recht, wichtig ist, wie wir unsere Beziehung gestalten.
- 5: Noll, Ingrid: Ladylike, Zürich: Diogenes 2006, S. 283  
»Ist es **okay**, wenn wir uns eine Odenwälder Wildrahmsuppe mit Waldpilzen bestellen?«



## **Grenzen: Semantische Blindheit**

Die großen deutschen Referenzkorpora sind bislang nicht semantisch annotiert. Man kann zwar nach (annotierten) Formeinheiten, aber nicht nach Wortbedeutungen suchen.

Auch die statistischen Auswertungen operieren auf formalen und nicht auf semantischen Einheiten.

Wer an semantischen Einheiten interessiert ist, muss unter Umständen viel Zeit für die intellektuelle Nachbereitung der Ergebnisse investieren.

Stollen|

Stollen<sup>1</sup>Stollen<sup>2</sup>

# Stollen, der

**Grammatik** Substantiv (Maskulinum) · Genitiv Singular: **Stollens** · Nominativ Plural: **Stollen**

**Worttrennung** Stol-len (computergeneriert)

**Wortbildung** mit ›Stollen‹ als Letztglied: ↗ [Christstollen](#) ... **2 weitere**

Bedeutung

Etymologie

Thesaurus

Typische Verbindungen

Verwendungsbeispiele


## Bedeutung

eWDG, 1976

- ▼ (besonders) süddeutsch, (besonders) österreichisch, (besonders) schweizerisch zu Weihnachten gebackener länglich-ovaler Hefekuchen von fester Beschaffenheit und mit vielen Zutaten

# Stollen, der

**Grammatik** Substantiv (Maskulinum) · Genitiv Singular: **Stollens** · Nominativ Plural: **Stollen**

**Aussprache** 

**Worttrennung** Stol-len (computergeneriert)

**Wortbildung** mit ›Stollen‹ als Erstglied: ↗ [Stollenbau](#) ... [3 weitere](#) · mit ›Stollen‹ als Letztglied: ↗ [Bruchstollen](#)  
... [8 weitere](#)

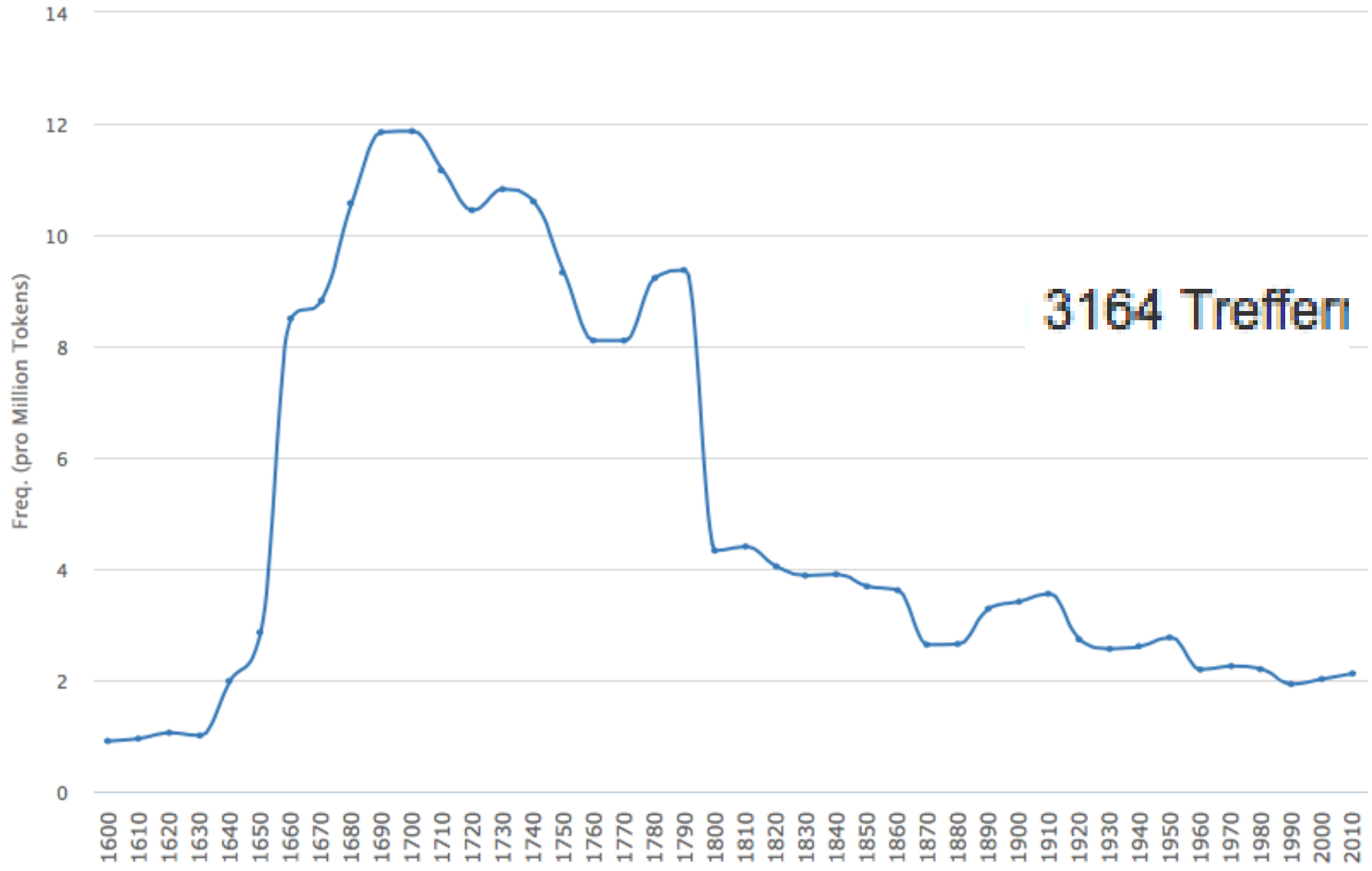
[Bedeutungen](#)[Etymologie](#)[Thesaurus](#)[Typische Verbindungen](#)[Verwendungsbeispiele](#)

## Bedeutungen

eWDG, 1976

- I. **Bergmannssprache** am Berghang meist leicht ansteigender, ins Gebirge getriebener, tunnelartiger Gang
- II. Zapfen am Hufeisen der Pferde oder an den Schuhen der Fußballspieler, der ein Ausgleiten verhindern soll
- III. **Literaturwissenschaft** eine der beiden gleichgebauten Strophen im Lied des Meistersanges

# Stollen - Verlaufskurve



3164 Treffer

# Typische Verbindungen

computergeneriert



*Sohle tritt auf mit Stollen*

Irgendwann im Laufe des Turniers riss dann noch ein **Stollen** aus der **Sohle**.

*Die Zeit, 01.10.2003, Nr. 40*

Und weil er offensichtlich die falschen **Stollen** unter den **Sohlen** hatte, rutschte er teilweise recht tollpatschig über den Rasen.

*Der Tagesspiegel, 09.11.2003*

Von der 12. **Sohle** des **Stollens**, wo die Wässer in einem Bassin zusammenlaufen, werden sie direkt in die Kurhäuser gepumpt.

*Die Zeit, 14.06.1991, Nr. 25*

Die lässige Handhabung der Berieselung des Kohlenstaubes hatte auf der Zeche "Gotthilf" in Gelsenkirchen einen Grubenbrand hervorgerufen, der sich mit unheimlicher Schnelligkeit verbreitet und mit seinen Rauchschwaden einen großen Teil der **Stollen** der letzten **Sohle** erfüllt hatte, daß mehr als hundert Bergarbeiter, von der Außenwelt abgeschnitten, dem sicheren Tode überliefert schienen.

*Stehr, Hermann: Der Heiligenhof, München: List 1952 [1918], S. 138*

#### 4. **Bergmannssprache** Gesamtheit der in einer horizontalen Ebene liegenden Strecken, **Stockwerk**

##### BEISPIELE:

auf der vierten **Sohle** ist ein Grubenbrand ausgebrochen

Der Korb geht nach **Sohle** sechs [KELLERM., *Tunnel*, 67]

# Linguistische Wikipedistik: Korpusbasierte Untersuchung des Sprachgebrauchs in der Wikipedia

## Deutsche Wikipedia-Korpora am IDS

<http://www1.ids-mannheim.de/kl/projekte/korpora/archiv/wp.html>

Korpus	Textwörter	Kopie vom
Wiki-A-de13	678.444.643	28.07.2013
Wiki-D-de13	272.654.185	
Wiki-A-de15	796.638.747	01.05.2015
Wiki-D-de15	309.897.027	
Wiki-A-de17	873.182.923	25.07.2017
Wiki-D-de17	349.075.823	



INSTITUT FÜR  
DEUTSCHE SPRACHE

Mitglied der Leibniz-Gemeinschaft



Wiki-A-de13: *Wikipedia Artikel Deutsch - Version 2013*

Wiki-D-de13: *Wikipedia Diskussionen Deutsch - Version 2013*

# Fallbeispiel: „Airlander“ in Wikipedia.de



WIKIPEDIA  
Die freie Enzyklopädie

Hauptseite  
Themenportale  
Von A bis Z  
Zufälliger Artikel

Mitmachen

Artikel verbessern  
Neuen Artikel anlegen  
Autorenportal  
Hilfe  
Letzte Änderungen  
Kontakt  
Spenden

Werkzeuge

Links auf diese Seite  
Änderungen an  
verlinkten Seiten  
Spezialseiten  
Permanenter Link  
Seiteninformationen  
Wikidata-Datenobjekt  
Artikel zitieren

Drucken/exportieren

Buch erstellen  
Als PDF  
herunterladen  
Druckversion

Artikel

Diskussion

Lesen

Bearbeiten

Quelltext bearbeiten

Versionsgeschichte

Suchen



## Airlander

Der **Airlander** ist ein Hybridluftschiff.

**Inhaltsverzeichnis** [Verbergen]

- 1 Airlander 10
- 2 Bisherige Flüge
- 3 Technische Daten
- 4 Airlander 50
- 5 Siehe auch
- 6 Weblinks
- 7 Einzelnachweise

### Airlander 10 [ Bearbeiten | Quelltext bearbeiten ]

Das Unternehmen Hybrid Air Vehicles entwickelte ab etwa 2010 als Subunternehmer des US-Rüstungskonzerns **Northrop Grumman**<sup>[1]</sup> im Rahmen des Projektes *Long Endurance Multi-intelligence Vehicle* (LEMV) den Prototyp **HAV 301**, der vom US-Verteidigungsministerium mit 60 Millionen US-Dollar finanziert wurde. Der HAV 301 absolvierte seinen Erstflug im August 2012 auf dem Luftschiffflugplatz in **Lakehurst** im US-Bundesstaat **New Jersey**.

Das Projekt sollte zu einem hoch fliegenden Aufklärungsluftschiff für die **US-Streitkräfte in Afghanistan** führen, wurde aber kurz nach dem Erstflug eingestellt.

Der britische Hersteller kaufte daraufhin den Prototyp für 301.000 US-Dollar zurück.<sup>[2]</sup>

Hybrid Air Vehicles wird von der deutschen **Linde AG** und deren britischer Tochtergesellschaft **BOC** gesponsert.

Airlander



Typ:	Hybridluftschiff
Entwurfsland:	<span><span><span></span></span><span> </span></span> Vereinigtes Königreich
Hersteller:	Hybrid Air Vehicles
Erstflug:	17. August 2016





## Northrop Grumman LEMV [\[ Quelltext bearbeiten \]](#)

---

Ja wie, ist das jetzt dasselbe wie [Northrop Grumman LEMV](#) ? -- [itu \(Disk\)](#) 00:52, 8. Mär. 2016 (CET)

Nein, wegen dem Bild meinst Du?--[MBurch \(Diskussion\)](#) 01:08, 8. Mär. 2016 (CET)

Unter anderem. -- [itu \(Disk\)](#) 15:36, 8. Mär. 2016 (CET)

Bild kann von mir aus weg, ansonsten ist es natürlich nicht derselbe.--[MBurch \(Diskussion\)](#) 19:38, 8. Mär. 2016 (CET)

Das Bild kann jedenfalls nicht einfach so dastehen als ob es den Lemmagegenstand abbilden würde, wenn dem gar nicht so ist; man kann das aber durchaus auch in der Bildunterschrift klarmachen. --

Ansonsten lese ich in [Northrop Grumman LEMV](#): „im Auftrag von Northrop Grumman durch eine britische Firma .. entwickelt und gebaut.“, sehe weiterhin irgendwie recht identische Grössen der Teile und im EN-interwiki-Artikel ist das Bild ebenfalls drin.

Ohne mich da jetzt reinstudiert zu haben ist das imho ganz schön viel Widersprüchliches. -- [itu \(Disk\)](#) 20:13, 8. Mär. 2016 (CET)

Dass ich *nicht verstanden werde* bedaure ich zutiefst, natürlich kann es auch sein dass es überhaupt nichts Widersprüchliches gibt und ich einfach nur komplett auf dem Schlauch stehe. -- [itu \(Disk\)](#) 20:18, 8. Mär. 2016 (CET)

Warum geht [Northrop Grumman LEMV](#) per interwiki (via Redirect [en:Long\\_Endurance\\_Multi-intelligence\\_Vehicle](#)) auf [en:Hybrid Air Vehicles HAV-3](#) was wiederum auf den umseitigen Artikel [Airlander](#) zurückinterwikit? Wie soll ich das nun verstehen? -- [itu \(Disk\)](#) 20:56, 8. Mär. 2016 (CET)

Schweigen. Derweil kommt der Airlander [aus der Halle](#)<sup>?</sup>. -- [itu \(Disk\)](#) 01:58, 26. Mär. 2016 (CET)

# Fallstudien auf der Basis von Wikipedia-Korpora

Untersuchungen zur Sprachvariation...

... im Kontext von Sprachverfallsdiskursen

z.B. Vergleich von Sprachmerkmalen in Wikipedia-Artikeln mit Sprachmerkmalen in Wikipedia-Diskussionsseiten (Storrer 2013a/b; Kupietz u.a. 2014; Storrer 2014, Storrer 2017)

Warum ist der Vergleich in diesem Kontext interessant?

Wiki-Artikel (in Wiki-A) sind Produkte des textorientierten Schreibens (monologisch, redigiert, starke Normorientierung)

Wiki-Diskussionen (in Wiki-D) sind Produkte des interaktionsorientierten Schreibens (situationsgebunden, dialogisch, schwache Normorientierung)

Wikipedia-Akteure schreiben auf beiden Seitentypen und wählen den jeweils passenden Schreibstil.

→ Hinweis darauf, dass das interaktionsorientierte Schreiben das textorientierte nicht ersetzt, sondern ergänzt.

# Wikipedia-Korpora zu verschiedenen Sprachen am IDS

## Umfang und Inhalt:

Artikel #Tokens	Artikel-Diskussionen #Tokens	Nutzer-Diskussionen # Tokens	
<b>Englisch</b> (wpe15, wde15, wue15)	2.403.943.177	1.270.217.981	2.698.338.998
<b>Französisch</b> (wpf15, wdf15, wuf15)	764.459.026	137.107.729	372.639.260
<b>Ungarisch</b> (wpu15, wdu15, wuu15)	117.987.947	8.293.799	26.215.158
<b>Norwegisch</b> (wpn15, wdn15, wun15)	99.014.144	5.314.362	32.481.331
<b>Spanisch</b> (wps15, wds15, wus15)	578.882.431	54.907.258	276.034.367
<b>Kroatisch</b> (wpk15, wdk15, wuk15)	46.641.724	2.480.966	18.731.167
<b>Italienisch</b> (wpi15, wdi15, wui15)	463.022.806	49.825.036	125.573.567
<b>Polnisch</b> (wpp15, wdp15, wup15)	298.207.197	16.558.557	64.126.136

Zur  
Aufbereitung  
: Margaretha  
& Längen  
(2014)

# Fallstudien auf der Basis von Wikipedia-Korpora

Sprachvergleichende Untersuchungen ...

... z.B. Vergleich von Wikipedia-Artikelseiten und Wikipedia-Diskussionen im Projekt EuroGr@mm: Typologisch und kontrastiv vergleichende grammatischen Erforschung und Beschreibung des Deutschen auf europäischer Ebene mit Vergleichssprachen: Deutsch, Französisch, Italienisch, Norwegisch, Polnisch, Ungarisch (vgl. Dalmas u.a. 2016).

Transmediale Untersuchungen ...

... .B. Vergleich des Gebrauchs von interaktiven Einheiten (Interjektionen / Responsiven) in Wikipedia-Diskussionen verschiedener Sprachversionen (Herzberg & Storrer 2017).

# „OK“ und „ok“ in Diskussionsseiten verschiedener Wikipedia-Sprachversionen (Version 2015)

Die Schreibformen „ok“, „OK“ sind in den Diskussionen vieler Sprachversionen der Wikipedia belegt.



**WIKIPEDIA**  
The Free Encyclopedia

122,1 pmw

69.906



**WIKIPÉDIA**  
L'encyclopédie libre

106,9 pmw

14.757



**WIKIPEDIA**  
Die freie Enzyklopädie

102,1 pmw

33.844



**WIKIPEDIA**  
L'enciclopedia libera

98,75 pmw

5.142



**WIKIPEDIA**  
La enciclopedia libre

29,27 pmw

1.671

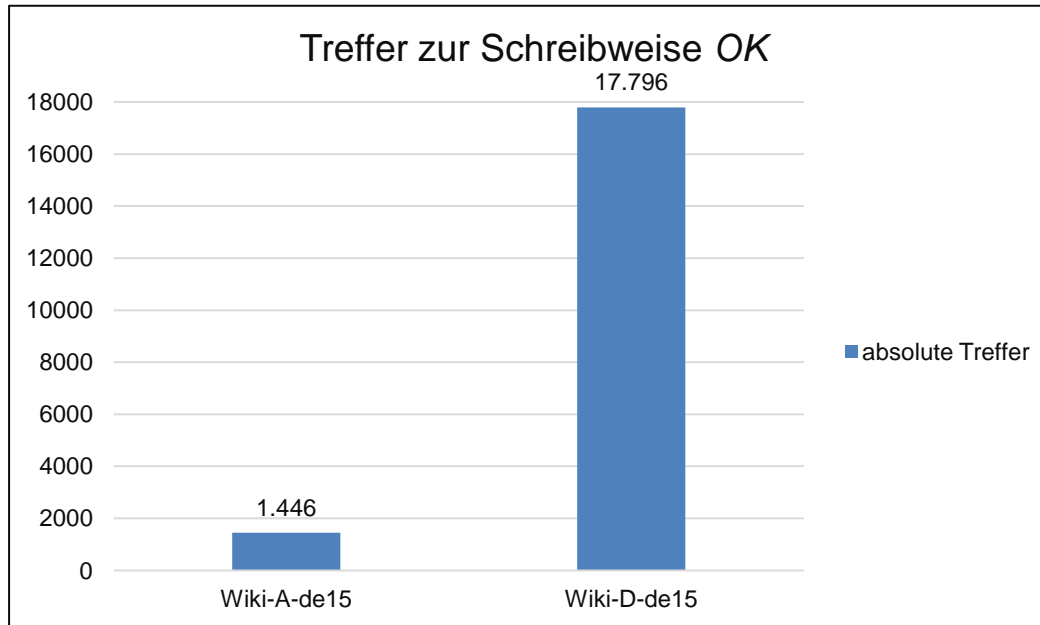
# Zum Vergleich: Frequenzen anderer deutscher Gesprächswörter auf den deutschen Diskussionsseiten



Wiki-D-de15

Frequenz	OK/ok	J/ja	N/nein	I/in Ordnung	E/einverstanden
pmw	102,1	2443,0	210,4	57,7	53,2
absolut	33.844	810.153	69.762	19.152	17.648

# Die Schreibform „OK“ in Wiki-A vs. Wiki-D



Frage: Handelt es sich hier überhaupt um Treffer für Schreibvarianten für das Lexem „OKAY“ oder um andere semantische Einheiten (z.B. Abkürzungen)?

→ Manuelle Überprüfung des Anteils von falsch Positiven (Pseudotreffern) in zwei Zufallsstichproben von 1000 Treffern pro Korpus

## Pseudotreffer (falsch Positive) zur Suchform OK

Wiki-A-de15

Artikelseite „Arbeitsmigranten in Kata“

Das Organisationskomitee (**OK**) der Weltmeisterschaft in Katar wehrte sich Anfang November 2013 gegen die Vorwürfe und sagte, es gebe auf den Baustellen erhebliche Fortschritte, die in den Medienberichten jedoch nicht erwähnt werden.

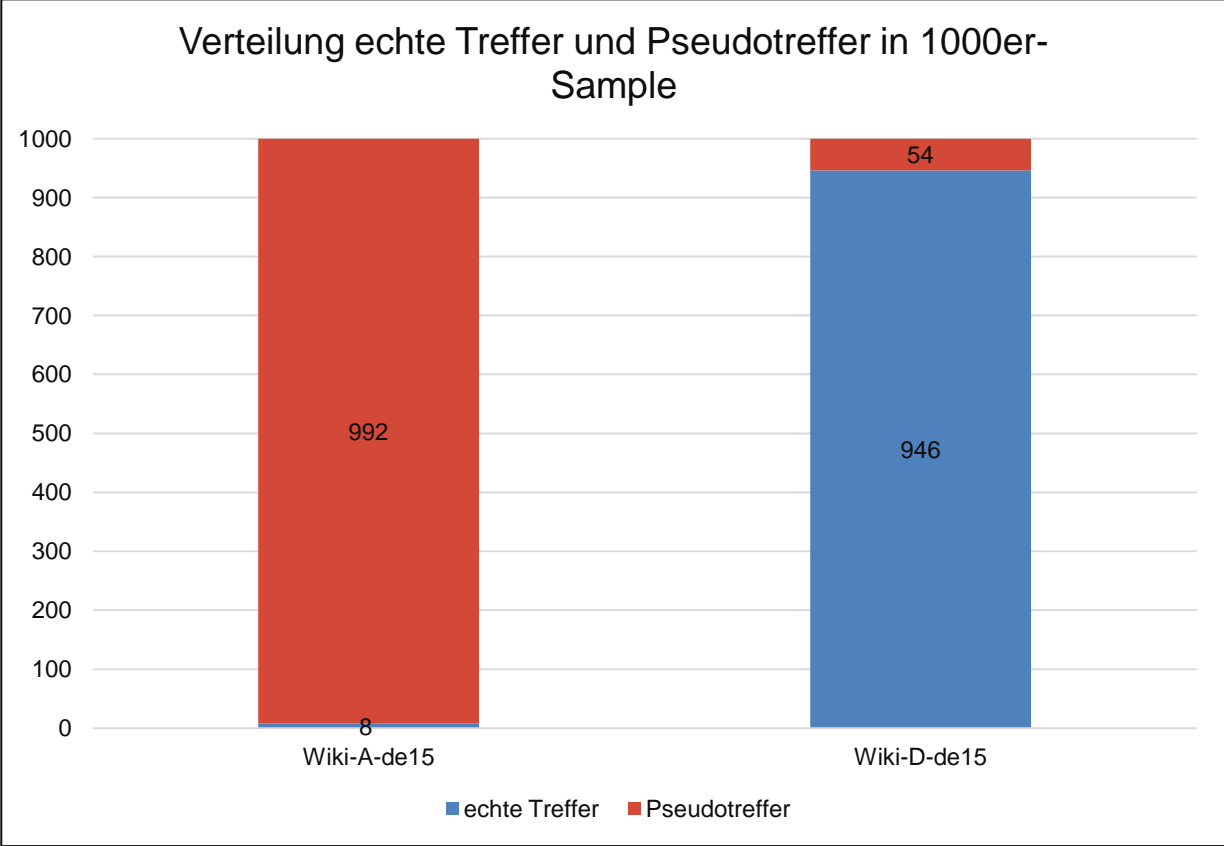
Wiki-D-de15

Diskussion zu: *Instrumentelle und operante Konditionierung*

Als weit ich den Textabschnitt verstehe unterscheidet sich die IK von der **OK** dadurch, dass die IK ein konditionierender Reiz ist, der auf ein zweckgerichtetes Verhalten wirkt; wohingegen die **OK** einen konditionierenden Reiz auf eine unbewusste, nicht zweckbestimmte Handlung ist.



# Die Schreibform „OK“ in Wiki-A vs. Wiki-D



	Wiki-A-de15	Wiki-D-de15
echte Treffer	8	946
falsch positive Treffer	992	54

## Vergleichsstudie Herzberg & Storrer (2017)

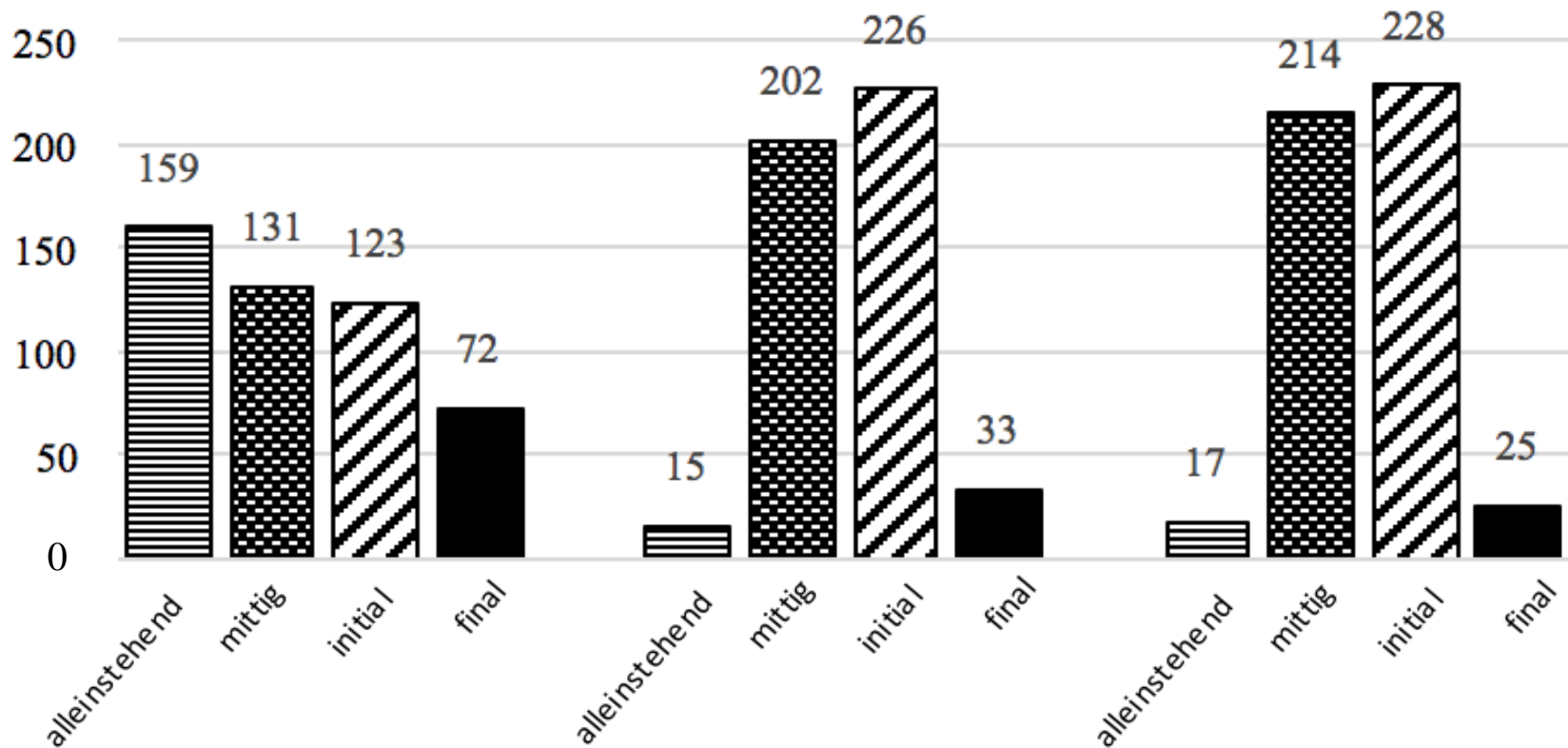
Korpus	Tokens
FOLK	1.952.159
Wiki-D-de15	309.897.027
Wiki-D-fr15	137.107.729

### Erkenntnisinteresse:

Entwickeln sich auf den Diskussionsseiten eigenständige Muster in den Verwendungsweisen von OKAY, die sich von den Mustern in der gesprochenen Sprache unterscheiden.

→ Vergleich von FOLK (gesprochene Sprache) und Wiki-D-de15 / Wiki-D-fr15

- Zufallsstichprobe von 500 Treffern pro Korpus
- Manuelle Aussonderung von Pseudotreffern
- Klassifikation nach vier Positionen: alleinstehend, initial, mittig, final.



FOLK

Wiki-D-de15

Wiki-D-fr15

## Vergleichsstudie Herzberg & Storrer (2017)

Korpus	Tokens
Wiki-D-de15	309.897.027
Wiki-D-fr15	137.107.729

### **Gibt es Unterschiede in der syntaktischen Integration zwischen der deutschen und der französischen Sprachversion?**

→ Vergleich der Anteile der syntaktisch integrierten Verwendungen (adverbial, prädikativ) vs. der Verwendung als interaktive Einheit (Responsiv, Rückmeldung, Diskursmarker etc.)

- Zufallsstichprobe von 500 Treffern pro Korpus
- Klassifikation nach zwei Funktionen: grammatisch integriert – interaktive Einheit.



Interaktive Einheit	61.2 %	58.9 %
syntaktisch integriert	37.5 %	10.3 %
andere	2.3 %	30.8 %

“Donc, OK pour moi.”

“OK pour la date de la mort.”

“L’introduction : OK.”

Die Daten mussten manuell annotiert werden, weil die POS-Tags nicht zwischen den Verwendungsweisen trennen.

## *Wunschzettel*

- Einheitliche Abfrageschnittstelle für verschiedene Korpora (für Vergleichsstudien).  
Möglichkeiten, Trefferlisten manuell zu bereinigen und die bereinigten Listen mit den Korpuswerkzeugen erneut auszuwerten.
- Verbesserung der linguistischen Basisannotation für Korpora mit hohem Anteil von standardabweichender Schreibung (Internetkommunikation, Gesprächstranskripte, historische Texte).
- Multimediale Korpora (= Text-Bild-Bezüge in digitalisierten Texten; multimodale Bezüge in genuin digitalen Texten (Websites, Social Media)).

# Literatur

Dalmas, M., C. Fabricius-Hansen & H. Schwinn (Hg.) (2016): Variation im europäischen Kontrast: Untersuchungen zum Satzanfang im Deutschen, Französischen, Norwegischen, Polnischen und Ungarischen. Berlin, New York: de Gruyter

Herzberg, Laura/Storrer, Angelika (2017): Investigating interaction signs across genres, modes and languages: The example of OKAY. In: Proceedings of the 5th Conference on CMC and Social Media Corpora for the Humanities (cmccorpora17), Bozen/Bolzano, S. 16-20. (doi: 10.5281/zenodo.1040875)

Kupietz, Marc; Belica, Cyril; Lungen, Harald; Perkuhn, Rainer (2014): Zwischen Empirie und Ästhetik – Ansätze zur korpuslinguistischen Untersuchung und Bewertung von Sprachwandel. In: Plewnia, A./ Witt, A. (eds.): Sprachverfall? Dynamik – Wandel – Variation. Jahrbuch des Instituts für Deutsche Sprache 2013, 149-170

Margaretha, Eliza/Harald Lungen (2014): Building linguistic corpora from Wikipedia articles and discussions. In: Journal of Language Technology and Computational Linguistics JLCL 29(2), 59–83.

Storrer, Angelika (2013a): Variation im deutschen Wortschatz am Beispiel der Streckverbgefüge. In: Deutsche Akademie für Sprache und Dichtung; Union der deutschen Akademien der Wissenschaften (Hrsg.): Reichtum und Armut der deutschen Sprache. Erster Bericht zur Lage der deutschen Sprache. Berlin/New York: de Gruyter, 171-209.

Storrer, Angelika (2013b): Sprachstil und Sprachvariation in sozialen Netzwerken. In: Frank-Job, Barbara/Mehler, Alexander/Sutter, Tilmann (Hg.): Die Dynamik sozialer und sprachlicher Netzwerke. Konzepte, Methoden und empirische Untersuchungen an Beispielen des WWW. Wiesbaden, S. 331-366.

Storrer, Angelika (2014): Sprachverfall durch internetbasierte Kommunikation? Linguistische Erklärungsansätze – empirische Befunde. In: Plewnia, A./ Witt, A. (eds.): Sprachverfall? Dynamik – Wandel – Variation. Jahrbuch des Instituts für Deutsche Sprache 2013, 171-196

Storrer, A. (2017): Grammaticale Variation in Gespräch, Text und internetbasierter Kommunikation. In Marek Konopka & Angelika Wöllstein (Hrsg.), *Grammatische Variation. Empirische Zugänge und theoretische Modellierung*. 105–125. Berlin, New York: de Gruyter.