

AntConc

2. Kollokationen, Keyness, POS-Tags

Kollokationen

Grundlegende Intuition

- Worte treten mit bestimmten anderen Worten zusammen auf und mit anderen nicht, obwohl dies semantisch auch möglich wäre.
 - Er sah aus wie ein geprügelter Hund.
 - Die Richterin verkündete das Urteil.
 - Die streunenden Katzen
 - Das war eine schöne Bescherung
- Die Verbindung kann unterschiedlich stark sein, manche Worte treten überhaupt nur zusammen aus, bei anderen handelt es sich um eine häufige Verbindung.

Definition

- „Die Identifikation von Kollokationen, verstanden als „nicht oder nur schwach idiomatische Wortverbindungen“ (Burger 2007), verlangt eine Abgrenzung in zwei Richtungen. Zum einen sind sie von Wortkombinationen abzugrenzen, deren Miteinandervorkommen in keiner Weise prädestiniert ist. Zum anderen sind sie von den übertragenen (idiomatischen) Wortverbindungen (Phrasemen) abzugrenzen, deren Bedeutung sich nicht wörtlich entschlüsseln lässt (bspw. „starker Tobak“, „kalter Kaffee“).“ (Kollokationenwörterbuch).
- Definition durch die **Korpuslinguistik**
„In corpus linguistics, a **collocation** is a sequence of words or terms that co-occur more often than would be expected by chance.“ (engl. Wikipedia 14.11.2017)

Kollokationen

- Wichtiges Instrument der Lexikographie
- Wichtig für den Spracherwerb, um idiomatische Sprachverwendung zu erlernen
- Wichtiges Instrument, um einen Einblick in die Semantik eines Worts zu bekommen

Hund

Hund¹ Hund²

Hund, der

Grammatik Substantiv (Maskulinum) · Genitiv Singular: **Hund(e)s** · Nominativ Plural: **Hunde**

Aussprache 

Wortbildung mit ›Hund‹ als Erstglied: ↗ [Hundchen](#) ... 58 weitere · mit ›Hund‹ als Letztglied: ↗ [Apportierhund](#) ... 64 weitere

Bedeutungen

Etymologie

Thesaurus

Typische Verbindungen

Verwendungsbeispiele

Bedeutungen

eWDG, 1969

- kleines oder mittelgroßes, zu den Raubtieren gehörendes Säugetier mit feinem Gehör und Geruchssinn, das wegen seiner Gelehrigkeit und Treue als Haustier gehalten, in unterschiedlichen Rassen gezüchtet und besonders als Wächter und bei der Jagd verwendet wird

BEISPIELE:

ein junger, großer, schöner, kluger, treuer, gehorsamer, rassereiner, herrenloser, wildernder, streunender, kläffender, rüddiger, toller, wilder, scharfer **Hund**

Vorsicht, bissiger **Hund**!

ein langhaariger, kurzhaariger **Hund**

... 19 weitere Beispiele

Ernst pfeift seinem **Hündchen**, das ihm das Halstuch zwischen den Zähnen bringt [SEGHERS, Siebtes Kreuz, 4,13]

▼ bildlich

BEISPIELE:

bei diesem Wetter jagt man keinen **Hund** vor die Tür, auf die Straße (= es ist sehr schlechtes Wetter)

das muss, kann (sogar) einen **Hund** jammern (= ist erbarmenswert) **salopp**

damit kann man keinen **Hund** hinter dem Ofen hervorlocken, vom Ofen locken (= damit kann man kein Interesse erwecken) **salopp**

von dem nimmt kein **Hund** ein Stück Brot mehr (= er ist allgemein verachtet) **salopp**

▼ übertragen

BEISPIELE:

jmdn. auf den **Hund** bringen (= jmdn. ruinieren)

auf den **Hund** kommen (= herunterkommen)

ganz auf dem **Hund** sein (= in schlechtem Zustand sein)

... 6 weitere Beispiele

- ohne Verkleinerungsform

salopp Mensch, besonders Mann

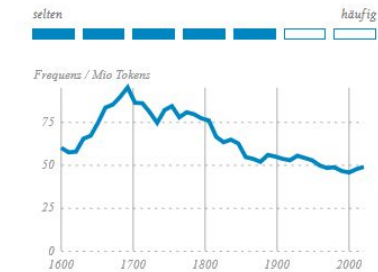
a) mit bestimmten Adjektiven

derb

BEISPIELE:

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

Worthäufigkeit



Ältere Wörterbücher

[Grimmsches Wörterbuch \('DWB\)](#) (3)

[Wörterbuch der deutschen Gegenwartssprache \(WDG\)](#) (2)

Korpustreffer

Referenzkorpora

[DWDS-Kernkorpus \(1900–1999\)](#) (7284)

[DWDS-Kernkorpus 21 \(2000–2010\)](#) (849)

[Deutsches Textarchiv \(1473–1927\)](#) (15538)

Zeitungskorpora

[Berliner Zeitung \(1994–2005\)](#) (14603)

[Tagesspiegel \(1996–2005\)](#) (7528)

[Die ZEIT \(1946–2016\)](#) (22105)

Spezialkorpora

[Referenz- und Zeitungskorpora \(aggregiert\)](#) (43360)

[Blogs](#) (7197)

[Polytechnisches Journal](#) (783)

[Filmuntertitel](#) (12608)

[Gesprochene Sprache](#) (89)

[DDR](#) (550)

Typische Verbindungen

DWDS-Wortprofil

computergeneriert



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

Verwendungsbeispiele

DWDS-Beispielextraktor

maschinell ausgesucht aus den DWDS-Korpora

Der **Hund** versuchte den Fuß zu erschnappen, aber das klappte nicht mehr.

Regener, Sven: Herr Lehmann, Frankfurt am Main: Eichborn AG 2006 [2001], S. 14

Als meine Kinder noch klein waren, hatten wir oft einen **Hund**.

Die Zeit, 30.03.2000, Nr. 14

Der **Hund**, der mit Bellen aufgehört hatte, jaulte jetzt vor sich hin.

Becker, Jürgen: Aus der Geschichte der Trennungen, Frankfurt a. M.: Suhrkamp 1999, S. 13

Der **Hund** hatte da eigentlich nichts zu suchen - aber wenn auch er zur Arbeit ging, so sei er willkommen.

Tudolsky, Kurt: Morgens um acht. In: Kurt Tudolsky, Werke - Briefe - Materialien, Berlin: Directmedia Publ. 2000 [1923]

Wortinformationen zu „Hund“ ... · Korpusuche nach „Hund“ ...

Suche im DWDS-Wortprofil

Lemma:

Hund

optional: Unterschiede zu

Vergleichswort

Anzahl:

min.

0

logDice:

min.

5

Sortierung:

logDice

Darstellung:

Überblick	logDice ↓↑	Freq. ↓↑
Katzen	9.2	3115 ^
bellen	7.9	975
gebissen	7.3	819
streunende	7.1	560
gefährlichen	7.0	1103
bunter	6.6	609
schlafende	6.6	397
armer	6.5	618
Pferde	6.3	557
gehetzt	6.3	356
Halter	6.3	325 v

hat Adjektivattribut	logDice ↓↑	Freq. ↓↑
streunende	9.4	560 ^
schlafende	8.8	397
gefährlichen	8.4	1023
bissigen	8.3	283
herrenlose	8.2	252
geprügelter	8.2	238
bellenden	8.0	210
bunter	8.0	608
armer	7.8	617
toten	7.8	481
pawlowschen	7.7	164 v

ist Akk./Dativ-Objekt von	logDice ↓↑	Freq. ↓↑
gehetzt	9.1	226 ^
ausführen	9.0	370
streicheln	8.3	155
anschaffen	7.9	106
anzuleinen	7.9	92
hervorlocken	7.7	81
füttern	7.7	101
beißen	7.4	79
ausführen	7.4	72
mitnehmen	7.3	176
wecken	7.2	198 v

ist in Präpositionalgruppe	logDice ↓↑	Freq. ↓↑
Leinenzwang für	8.5	170 ^
Schwanz mit	8.3	143
wedelt mit	8.2	147
Spaziergang mit	8.0	143
gebissen von	7.9	159
Dame mit	7.6	139
wackelt mit	7.0	63
Halten von	6.8	50
angefallen von	6.7	82
Besitzer von	6.6	53
Halter von	6.6	44 v

hat Genitivattribut	logDice ↓↑	Freq. ↓↑
Rassen	10.4	151 ^
Spaziergängers	9.8	33
Nachbarn	8.5	85
Nachbarin	8.2	13
Kategorie	8.1	40
Spaziergängerin	8.1	9
Schäfers	7.9	8
Försters	7.7	7
Bekannten	7.2	9
Schwiegermutter	7.0	5
Größe	6.9	17 v

in Koordination mit	logDice ↓↑	Freq. ↓↑
Katzen	12.2	3060 ^
Pferde	9.6	537
Katz	9.4	315
Herrchen	8.9	229
Vögel	8.7	263
Schweine	8.6	256
Kaninchen	8.5	186
Hühner	8.4	207
Katzen	8.4	149
Affen	8.3	163
Ratten	8.2	159 v

hat Prädikativ	logDice ↓↑	Freq. ↓↑
aggressiv	9.0	56 ^
Hund	8.9	32
Tier	8.6	28
Gefährte	8.2	12
Haustiere	8.2	12
Begleiter	8.2	15
angeleint	8.2	11
eingeschläfert	8.2	11
Tod	8.1	23
Kampfhunde	8.0	10
Freund	7.8	76 v

hat vergl. Wortgruppe	logDice ↓↑	Freq. ↓↑
als Haustier	10.8	24 ^
als Zugtier	10.4	7
als Gefährte	10.3	6
als Begleiter	8.9	9
als Statussymbol	8.4	5
als Waffe	7.5	18
als Helfer	6.9	5
als Tier	6.6	6
als Freund	6.0	11
als Menschen	4.2	13
als Kinder	3.0	5 v

ist Subjekt von	logDice ↓↑	Freq. ↓↑
angetaen	7.1	119 ^
heulen	7.1	88
sprang	7.1	182
schnüffeln	7.1	72
wedelt	7.0	72
verbissen	6.9	66
anbellt	6.9	63
streunen	6.9	62
angeleint	6.7	54
pinkeln	6.5	51
knurrt	6.5	51 v

hat Präpositionalgruppe	logDice ↓↑	Freq. ↓↑
an Leine	9.8	197 ^
an Leine	9.4	163

ist Genitivattribut von	logDice ↓↑	Freq. ↓↑
Bellen	10.1	111 ^
Halter	9.9	112

ist Prädikativ von	logDice ↓↑	Freq. ↓↑
Hund	8.9	32 ^
Begleiter	8.1	8

KOLLOKATIONENWOERTERBUCH

Feste Wortverbindungen des Deutschen. Kollokationenwörterbuch für den Alltag

Hund

☐ Bsp./Infos ☐ ganze Form

[Wörterbuch](#) [Projekt](#) [Aktivitäten](#) [Kontakt](#)

Hund

hundefrei

Hundert

hundertjährig

hundertmal

hundertprozentig

geschunden

Jahrhundert

jahrhundertealt

Hund m

ADJEKTIVE/ADVERBIEN

brav • lieb • treu • bellend • bissig • kläffend • angeleint • frei laufend •
scharf • tollwütig • streunend • herrenlos

ausgewachsen • abgerichtet •
nass • reinrassig • stubenrein

VERBEN

halten • besitzen • haben • halten •
• züchten

hetzen • hetzen • jagen • zlehen

erziehen • abrichten • erziehen •
gehören • einem Hund pfeifen •

rufen • pfeifen • rufen •
zurückrufen •

festhalten • anleinen • festhalten •
kastrieren • einschläfern •

apportieren • bewachen •
erlauben • legen • streicheln •
wedeln • zulaufen

NOMEN

Charakter • Erziehung • Kette • Leine • Tierhelm • Zucht • Maulkorb • Stammbaum

Herrin Herr • Frauchen •
Herrchen • Katze

ZUSAMMENSETZUNGEN

Suchhund • Spürhund •
Wachhund • Lawinhund •
Jagdhund • Kampfhund •
Hütehund • Rettungshund •
Polizeihund • Herdenhund •
Schlittenhund •

Rassehund • Mischlingshund •
Straßenhund • Schäferhund •
Hundefutter • Hundeleine •
Hundemarke • Hundehütte •
Hundehaar • Hundesteuer

Hundehaltung • Hundezucht •
Hunderasse • Hundewelpe •
Hundegebell •
Hundebesitzer_in •
Hundehalter_in •
Hundeführer_in •

Verwendete Maße I: Mutual Information (MI)

- Das Maß ‚gemeinsame Information‘ drückt aus, wieviel man über das Vorkommen von des Wortes y weiß, wenn man weiß, dass das Wort x vorkommt. Kommt das Wort y ganz unabhängig vom Wort x vor, dann gibt es keine gemeinsame Information, dann ist die MI 0. Kommen die Worte dagegen häufiger gemeinsam vor, dann wird der Wert höher sein.

Verwendete Maße II: t-test

- Der t-test wird in der Statistik verwendet, um die Wahrscheinlichkeit zu berechnen, dass sich der Wert einer Messung ergibt, wenn die Nullhypothese richtig ist, wobei in unserem Fall die Nullhypothese die Unabhängigkeit der beiden Wörter ist. Wenn der t-test signifikant ist, dann bedeutet das in diesem Fall, dass die beiden Wörter häufiger gemeinsam vorkommen, als es in einer ganz zufälligen Verteilung wahrscheinlich wäre. Wird aber nur zum Ranking der Ergebnisse verwendet, da fast immer signifikant.

Berechnung von Kollokationen

- Gezählt wird die Häufigkeit von Wörtern.
- f_x = Häufigkeit des Wortes x im Korpus.
- f_y = Häufigkeit des Wortes y im Korpus.
- f_{xy} = Häufigkeit des gemeinsamen Auftauchens der Worte x und y im Text. „gemeinsames Auftauchen“ bedeutet, dass höchstens ein festgelegter Abstand, zumeist 2 oder 3 Worte, zwischen den Worten liegen darf. Man spricht häufig auch von einem ‚Fenster‘ :

Die alte RichterIn verkündete mit ruhiger Stimme das Urteil.



Fenstergröße von 3 links und 3 rechts

Kollokationen in AntConc

Kollokationen in AntConc

- Man sucht immer Kollokationen zu einem Wort / Wortgruppe / String
- Für die Angabe des Worts kann man all das verwenden, was wir oben bei der ‚Suche‘ kennengelernt haben
- Man kann festlegen, wie groß das Fenster um das Wort ist, in dem die Kollokationen ermittelt werden

Corpus Files

Alexis_Willibald-Herr-von-Sacken.txt
 Andolt_Ernst-Eine-Nacht.txt
 Annm_Achim-von-Der-tolle_invalide_auf_dem_Fort_Ratoneau.txt
 Auerbach_Berthold-Die_Geschichte_des_Diethelm_von_Buchenberg.txt
 Berthold_Franz-Irwinch-Fritze.txt
 Brentano_Clemens-Geschichte_vom_braven_Kasperl_und_dem_schoenen_Annerl
 Chamisso_Adelbert-von-Peter_Schleimhl_s_wundersame_Geschichte.txt
 Dindlage_Emma_v.-Der_Striebstast.txt
 Droste_Huelshoff_Annette-von-Die_Judenbuche.txt
 Eichendorff_Joseph-von-Die_Gluecksritter.txt
 Ense_Karl_August_Varnhagen_von-Reiz_und_Liebe.txt
 Frey_Jacob-Das_erfuellte_Versprechen.txt
 Gall_Louise-von-Eine_fromme_Lüge.txt
 Gerstaecker_Friedrich-Germelshausen.txt
 Gluemer_Claire-von-Reich_zu_reich_und_arm_zu_arm.txt
 Goethe_Wolfgang-von-Die_neue_Melusine.txt
 Goldammer_Leo-Auf_Wiedersehen!.txt
 Goldammer_Leo-Eine_Hochzeitsnacht.txt
 Gotthelf_Jeremias-Der_Notar_in_der_Falle.txt
 Gotthelf_Jeremias-Kurt_von_Koppigen.txt
 Grillparzer_Franz-Der_arme_Spielmann.txt
 Grimm_Herman-Das_Kind.txt
 Grosse_Julius-Vetter_Isidor.txt
 Hacklaender_Friedrich_Wilhelm-Zwei_Naechte.txt
 Halm_Friedrich-Die_Marzipan_Lise.txt
 Hartmann_Moritz-Das_Schloss_im_Gebirge.txt
 Hauff_Wilhelm-Phantasien_im_Bremer_Ratskeller.txt
 Heyden_Friedrich-von-Der_graue_John.txt
 Heyse_Paul-Der_Weinhuetter_von_Meran.txt
 Hoefer_Edmund-Rolof_der_Rekrut.txt
 Hoffmann_E.T.A.-Das_Fraelein_von_Scuderi.txt
 Holtei_Karl_von-s_Muhme_Leutnant_Saloppel.txt
 Horner_Heinrich-Der_Säugling.txt
 Immermann_Karl-Der_Carneval_und_die_Somnambule.txt
 Kaehler_Ludwig-August-Die_drei_Schwestern.txt
 Keller_Gottfried-Romeo_und_Julia_auf_dem_Dorfe.txt
 Kinkel_Gottfried-Margret.txt
 Kinkel_Johanna-Musikalische_Orthodoxie.txt
 Kleist_Heinrich-von-Die_Verlobung_von_St_Domingo.txt
 Kompert_Leopold-Eine_Verlorene.txt
 Kopisch_August-Der_Trauerer.txt
 Kopisch_August-Ein_Karnevalsfest_auf_Ischia.txt
 Kruse_Laurids-Nordische_Freundschaft.txt
 Kuernberger_Ferdinand-Der_Drache.txt
 Kugler_Franz-Die_Incantada.txt
 Kurz_Hermann-Die_beiden_Tubus.txt
 Lewald_Fanny-Die_Tante.txt
 Lohmann_Friederike-Die_Entscheidung_bei_Hochkirch.txt
 Lorm_Hieronymus-Ein_adeliges_Fraelein.txt
 Ludwig_Julie-Das_Gericht_im_Valde.txt
 Meißner_Alfred-Der_Müller_vom_Höft.txt
 Meyr_Melchior-Der_Sieg_des_Schwachen.txt
 Moenike_Eduard-Mozart_auf_der_Reise_nach_Prag.txt
 Muegge_Theodor-Am_Malanger_Fjord.txt
 Mueller_Wilhelm-Debora.txt
 Pichler_Adolf-Der_Fluechtling.txt
 Raabe_Wilhelm-Das_letzte_Recht.txt
 Reich_Moritz-Mammon_im_Gebirge.txt
 Riehl_Wilhelm_Heinrich-Joerg_Muckenhuber.txt
 Roquette_Otto-Die_Schlangenkönigin.txt
 Rumohr_Karl_Friedrich-Der_letzte_Savello.txt
 Sacher_Masoch_Leopold-Don_Juan_von_Kolomea.txt
 Schefer_Leopold-Die_Duevecke_oder_die_Leiden_einer_Koenigin.txt
 Scheffel_Joseph_Victor_von_Hugideo.txt
 Schmid_Hermann-Mohrenfranzel.txt
 Schreyvogel_Joseph-Samuel_Brinks_letzte_Liebesgeschichte.txt
 Schuecking_Levin-Die_Schwester.txt
 Spindler_Karl-Die_Engel_Ehe.txt
 Sternberg_Alexander_von-Scholastika.txt
 Stifter_Adalbert-Brigitta.txt
 Storm_Theodor-Eine_Malerarbeit.txt
 Tesche_Walter-Der_Enten_Piet.txt
 Tieck_Ludwig-Das_Lebens_Ueberfluss.txt
 Tieck_Ludwig-Die_Gemaelde.txt
 Trautmann_Julius-von-Der_Gehirnferretur.txt

Total No.
86

Files Processed

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Total No. of Collocate Types: 986 Total No. of Collocate Tokens: 2430

Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	2	1	1	14.50649	höllenhunde
2	2	1	1	14.50649	gebissenen
3	1	1	0	13.50649	zusammengezählt
4	1	0	1	13.50649	zahllos
5	1	1	0	13.50649	wolfszahn
6	1	1	0	13.50649	wegfressen
7	1	1	0	13.50649	verwöhnte
8	1	0	1	13.50649	verwechslung
9	1	0	1	13.50649	versatz
10	1	1	0	13.50649	unbeholfener
11	1	0	1	13.50649	umwarfen
12	1	1	0	13.50649	umgerannt
13	2	1	1	13.50649	tractirt
14	1	1	0	13.50649	tauber
15	1	1	0	13.50649	tafelten
16	1	0	1	13.50649	strohbohrer
17	1	1	0	13.50649	stachelten
18	1	0	1	13.50649	stacheligem
19	1	1	0	13.50649	spickt
20	1	1	0	13.50649	schöcherbeth
21	1	1	0	13.50649	schnaufens
22	3	3	0	13.50649	räudigen
23	1	1	0	13.50649	respekte
24	1	1	0	13.50649	rathsnachtwächter
25	1	0	1	13.50649	rascherem
26	1	0	1	13.50649	raisonniren
27	1	1	0	13.50649	quin
28	1	0	1	13.50649	persern
29	1	1	0	13.50649	ochsenhüten
30	1	1	0	13.50649	nägelspuren
31	1	1	0	13.50649	niedergeschossen
32	1	1	0	13.50649	miauten
33	1	1	0	13.50649	meuterer
34	1	1	0	13.50649	messingringen
35	1	0	1	13.50649	menschengesichtern
36	1	0	1	13.50649	markdurchdringender
37	1	1	0	13.50649	machtvolle
38	1	0	1	13.50649	leite
39	1	1	0	13.50649	lappiger
40	1	0	1	13.50649	klaffend
41	1	0	1	13.50649	jagdgehege
42	1	0	1	13.50649	hudeln
43	2	0	2	13.50649	hinausjagen
44	1	1	0	13.50649	herzdurchschneidender
45	1	0	1	13.50649	herumtrampelt
46	1	0	1	13.50649	herangesprungen
47	1	1	0	13.50649	gekniert
48	1	0	1	13.50649	gehörnte
49	1	0	1	13.50649	gampel

Search Term ☒ Words ☐ Case ☐ Regex

Hund|Hunde|Hündin

Advanced

Start

Stop

Sort

Sort by ☐ Invert Order

Sort by Stat

Window Span ☐ Same

From... 5L To... 5R

Min. Collocate Frequency

1

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Total No. of Collocate Types: 986 Total No. of Collocate Tokens: 2430

Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	2	1	1	14.50649	höllenhunde
2	2	1	1	14.50649	gebissenen

Gesamthäufigkeit.
Wichtiges Korrektiv

Fenster von 5
Worten links und
rechts um ‚Hund‘

Search Term ☒ Words ☐ Case ☐ Regex

Hund|Hunde|Hündin

Advanced

Start

Stop

Sort

Sort by ☐ Invert Order

Sort by Stat

Window Span ☐ Same

From... 5L To... 5R

Min. Collocate Frequency

1

Sortiert nach Häufigkeit

AntConc 3.4.4w (Windows) 2014

File Global Settings Tool Preferences Help

Corpus Files

Alexis_Willibald-Herr_von_Sacken.txt
Andolt_Ernst-Eine_Nacht.txt
Arnim_Achim_von-Der_tolle_invalide_auf_dem_Fort_Ratoneau.txt
Auerbach_Berthold-Die_Geschichte_des_Diethelm_von_Buchenberg.txt
Berthold_Franz-Irrwisch-Fritze.txt
Brentano_Clemens-Geschichte_vom_braven_Kasperl_und_dem_schoenen_Annerl
Chamisso_Adelbert_von-Peter_Schlemihl_s_wundersame_Geschichte.txt
Dincklage_Emma_v.-Der_Striethast.txt
Droste_Huelshoff_Annette_von-Die_Judenbuche.txt
Eichendorff_Joseph_von-Die_Gluecksritter.txt
Ense_Karl_August_Varnhagen_von-Reiz_und_Liebe.txt
Frey_Jacob-Das_erfuellte_Versprechen.txt
Gall_Louise_von-Eine_fromme_Lüge.txt
Gerstaeker_Friedrich-Germelshausen.txt

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Total No. of Collocate Types: 227

Total No. of Collocate Tokens: 1348

Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	86	57	29	4.66417	der
2	82	56	26	4.57481	die
3	73	33	40	3.94234	und
4	36	32	4	5.41095	wie
5	34	27	7	4.84513	ein
6	32	20	12	4.28023	den
7	25	10	15	4.06861	nicht
8	25	8	17	3.75234	ich

Informationsarme
Funktionswörter

Spieler_von-Die_Enger_Engel.txt
Sternberg_Alexander_von-Scholastika.txt
Stifter_Adalbert-Brigitta.txt
Storm_Theodor-Eine_Malerarbeit.txt
Teschke_Walter-Der_Enten_Piet.txt
Tieck_Ludwig-Des_Lebens_Ueberfluss.txt
Tieck_Ludwig-Die_Gemaelde.txt
Traub_Julius_von-Der_Der_Gebirgsfarrer.txt

Total No.

86

Files Processed

47	6	0	6	3.26968	des
48	6	1	5	3.46803	auch
49	5	2	3	3.74460	vor

Search Term ☒ Words ☐ Case ☐ Regex

Hund|Hunde|Hündin

Advanced

Start

Stop

Sort

Sort by ☐ Invert Order

Sort by Freq

Window Span ☐ Same

From...

4L

To...

4R

Min. Collocate Frequency

2

keyness

Keyness

- Ermittelt die Worte, in denen sich eine Textgruppe von einer anderen unterscheidet.
- In AntConc: Die angezeigte Liste zeigt, welche Worte deutlich häufiger in dem untersuchten Korpus (im Vergleich zum Referenzkorpus) auftauchen.
- Zwei Maße:
 - Log likelihood
 - Chi Squared
 - Beide berechnen, ob der Unterschied zwischen den zu erwartenden Werten und den realen Werten so groß ist, dass wahrscheinlich zwei Verteilungen vorliegen
- Praktischer Umgang: Nur die ersten Worte sind aussagekräftig

Vorgehensweise

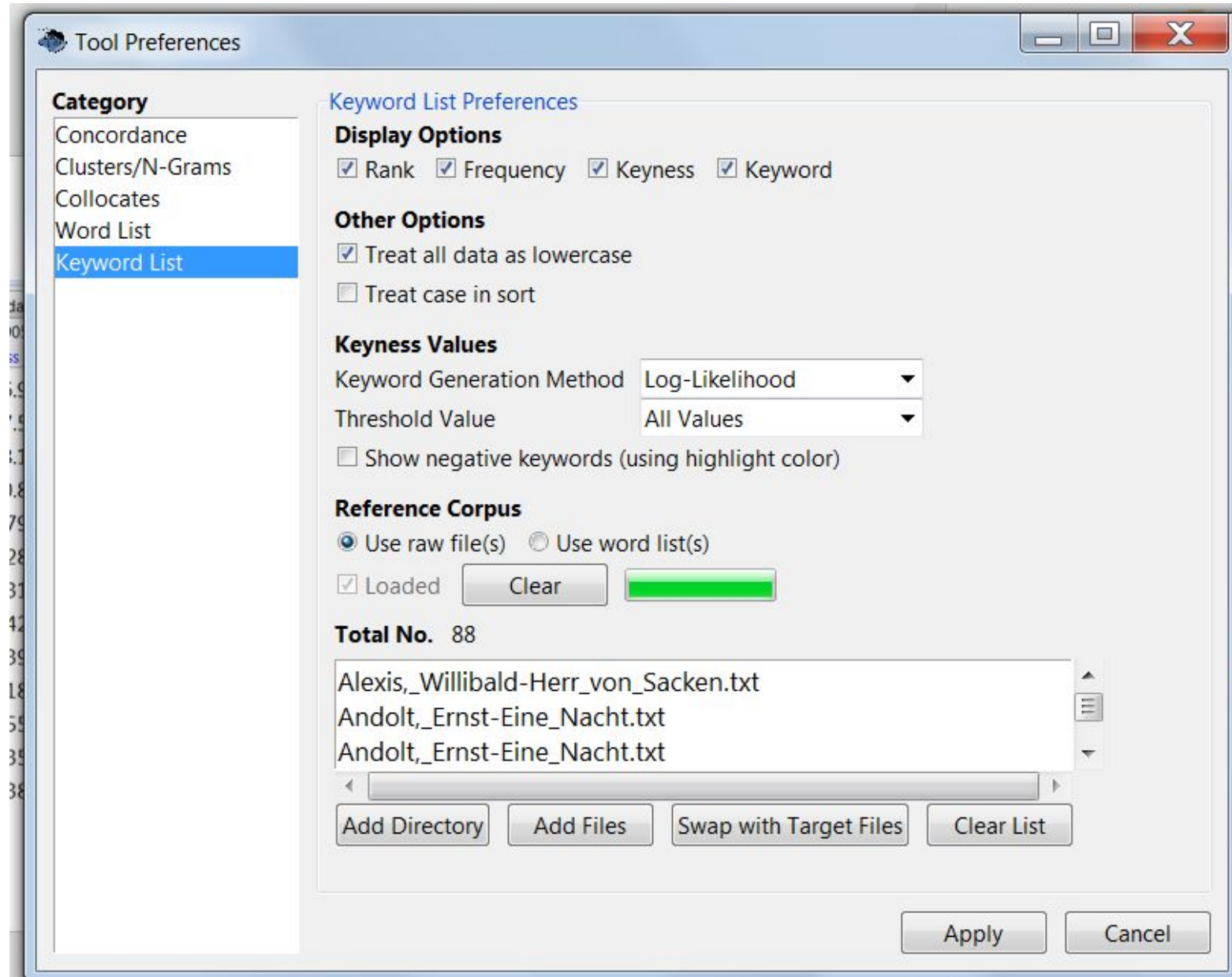
- Festlegung des 1. Korpus
- Festlegung des 2. Korpus
- Wahl eines Keyness-Maßes
- Analyse anstoßen
- Ergebnis der Analyse - die Wortliste - untersuchen

Vorgehensweise in AntConc

- Festlegung des 1. Korpus: Wie gehabt im Hauptfenster
 - Voraussetzung: Kopien der Texte in einem Verzeichnis
- Das 2. Korpus: Menüpunkt „Tool Preferences“ -> „Keyword List“
 - Voraussetzung: Kopien der Texte in einem Verzeichnis
- Keyness-Maß: Menüpunkt „Tool Preferences“ -> „Keyword List“
- Analyse anstoßen: „Start“
- Ergebnis der Analyse - die Wortliste – im Hauptfenster
 - Optional: Sortierung

Keyness in AntConc

- Fokuskorpus auswählen wie gewohnt
- Reference Korpus auswählen:
Tool Preferences-> Keyword List
- Einstellungen:
- Use raw files -> Load -> Apply



Vergleich weibl. und männl. Autoren im Novellenschatz

Types Before Cut: 19181			Types After Cut: 19181
Rank	Freq	Keyness (LL)	Keyword
1	351	741.991	leonie
2	4892	672.147	sie
3	174	339.972	louis
4	141	287.014	marie
5	118	238.748	ida
6	112	236.761	claudine
7	1532	232.515	ihr
8	140	231.011	gräfin
9	109	230.419	bernhard
10	109	230.419	otto
11	101	213.507	marquis
12	228	210.821	graf
13	96	202.938	liesbeth
14	87	183.912	klemenz
15	83	175.457	henriot
16	76	160.659	françois
17	75	158.545	caroline
18	74	156.431	therese

Vergleich von Dramen und Novellen

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Types Before Cut: 19055		Types After Cut: 17795		Search Hits: 0		
Rank	Freq	Keyness	Keyword			
1	943	4396.907	diethelm			
2	289	1347.514	fränz			
3	249	1138.103	lieschen			
4	284	1070.811	fritz			
5	247	866.798	munde			
6	180	839.282	sacken			
7	177	749.313	fragte			
8	195	746.421	fuhr			
9	153	713.390	medard			
10	268	601.185	ging			
11	134	596.550	baron			
12	210	560.351	wagen			
13	291	531.383	sah			

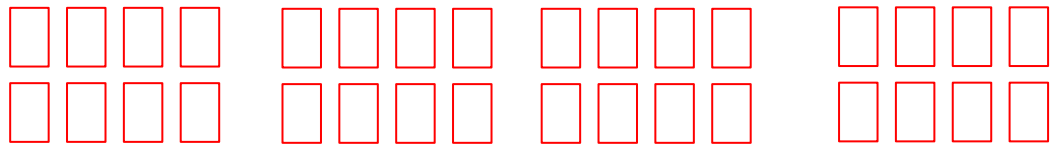
Vergleich männl. und weibl. Autoren im Novellenschatz

Rank	Freq	Keyness	Keyword
1	279	322.876	graf
2	140	235.716	gräfin
3	101	217.034	marquis
4	397	135.566	frau
5	58	124.633	schlichting
6	78	110.225	frug
7	453	106.828	kind
8	184	106.760	liebe
9	497	103.634	mann
10	40	85.954	sohling
11	436	83.319	vater
12	155	80.419	o
13	37	79.508	pistor
14	59	78.032	onkel

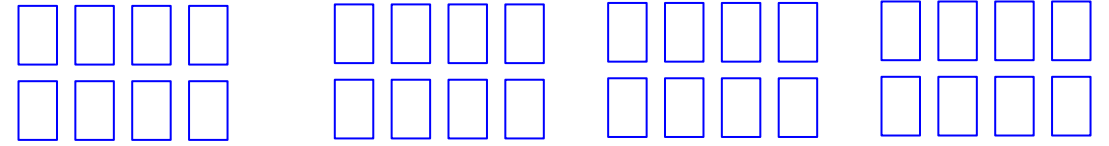
Keyword List Results 2				—	□	×
Types Before Cut: 19618		Types After Cut: 19618		Search Hits: 0		
Rank	Freq	Keyness (LL)	Keyword			
1	714	88.558	herr			
2	90	75.201	sacken			
3	70	58.489	fischele			
4	69	57.654	parzik			
5	404	55.907	nacht			
6	86	52.067	doctor			
7	62	51.805	silberburg			
8	60	50.134	klarinet			
9	226	47.944	pferd			
10	56	46.791	francoeur			
11	55	45.956	neger			
12	83	45.497	müller			
13	54	45.120	kasper			
14	54	45.120	koppigen			
15	63	44.487	kloster			
16	51	42.614	reinbacher			
17	50	41.778	beute			

Burrow's Zeta

Dispersionsmaß zum Auffinden distinktiver Wörter für Textgruppen





Western



Liebesroman

Burrow's Zeta

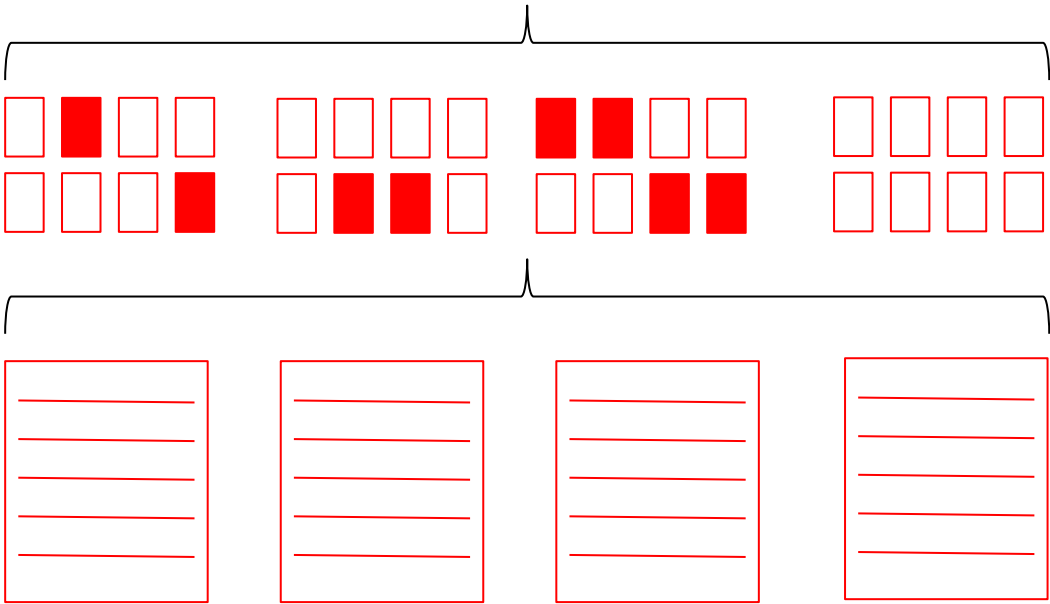
Berechnung von Zeta Wert für
"Colt"

Wort in Segment 
Wort in nicht Segment 

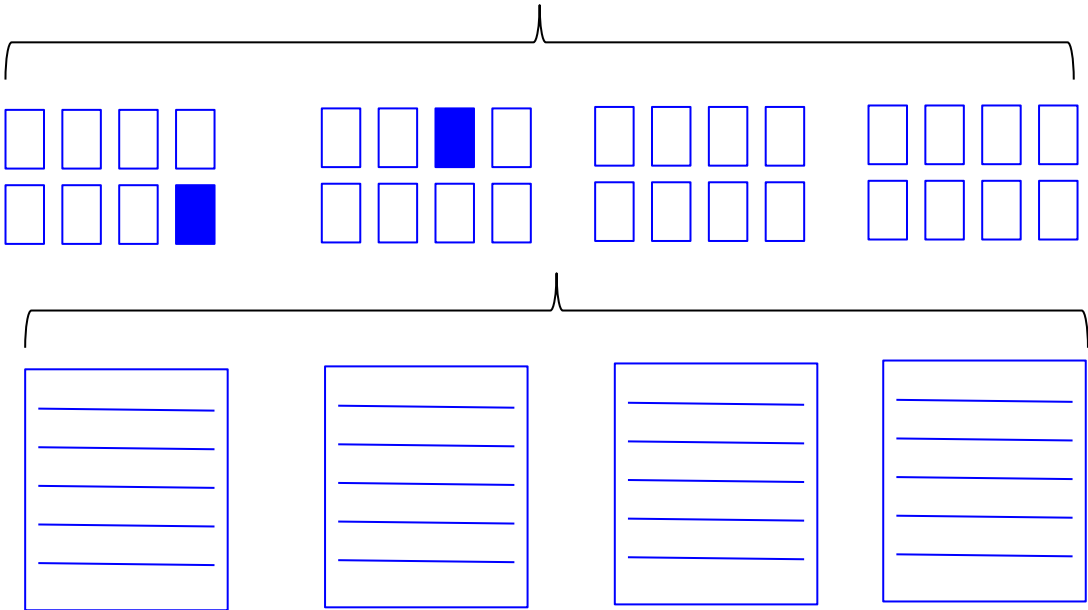
Zeta:
 $0.25 - 0.06 = 0.19$

DP: 8/32: 0.25

DP: 2/32: 0.06



Western



Liebesroman

Heimat	Kriegs	Krimi	Liebes	SciFi	Hochlit
madl	leutnant	streifenwagen	dienerschaft	galaxis	klo
bissel	mg	ford	gnädigen	raumschiff	it
brotzeit	munition	field	diwan	planet	me
tonis	russen	officer	gnädiges	universums	hitler
bös	russischen	schalldämpfer	anerbieten	schleuse	texte
obstler	deutscher	handschellen	unbeschreiblichen	weltraum	for
förster	einschläge	dienstwaffe	vornehmer	hangar	weltkrieg
ausschaut	flugzeuge	detective	gottlob	schutzschirm	on
gell	oberst	brooklyn	destille	raumschiffs	andauernd
leut	funker	notebook	teetisch	raumfahrer	wischt
trenker	meldet	inspektoren	liebenswürdigkeit	jahrtausenden	juden
bergwald	flanke	ermittlung	mancherlei	terra	russland
bursch	lastwagen	mafia	reizendes	lichtjahre	cola
bergführer	feindes	plaza	umzukleiden	humanoiden	wörter
feschen	pistole	ganoven	umkleiden	geortet	präsidenten
gerad	ne	bewußte	namenlos	unsterblichen	what
bergtour	flieger	datenbanken	frohen	projektion	christus

Keyness Experimente

- Keyness ermöglicht Textgruppenvergleich
- Textgruppen werden vorher festgelegt
- Ermittelt werden die markanten Wörter
- Die Gruppen müssen vorher festgelegt werden
- Problem: Bestätigt vorhandenes Wissen. Kaum Widerlegung

Part-of-Speech Tagging

Filtern mit Part-of-Speech (POS) Tags

- POS-Tags: Adj (Adjektiv), NN (normales Nomen), NE (Eigennamen), ART (Artikel), \$. (Satzbeendende Interpunktion) usw.
- Referenz fürs Deutsche: Stuttgart-Tübinger Tagset (STTS)
<http://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/TagSets/stts-table.html>



Sie sind hier: [Institut für Maschinelle Sprachverarbeitung](#) » [Forschung](#) » [Ressourcen](#) » [Lexika](#) »

Institut

[Arbeitsgruppen](#)
[Mitarbeiter](#)
[Stellenangebote](#)
[Veranstaltungen](#)
[Kontakt](#)
[Wiki \(intern\)](#)

Studium

[Interessierte](#)
[Studierende](#)
[Austauschprogramme](#)
[Fachgruppe](#)

Forschung

[Projekte](#)
[SFB 732](#)
[Publikationen](#)
[Ressourcen](#)
[Demos](#)

STTS Tag Table (1995/1999)

POS	DESCRIPTION	EXAMPLES
ADJA	attributives Adjektiv	[das] große [Haus]
ADJD	adverbiales oder prädikatives Adjektiv	[er fährt] schnell, [er ist] schnell
ADV	Adverb	schon, bald, doch
APPR	Präposition; Zirkumposition links	in [der Stadt], ohne [mich]
APPRART	Präposition mit Artikel	im [Haus], zur [Sache]
APPO	Postposition	[ihm] zufolge, [der Sache] wegen
APZR	Zirkumposition rechts	[von jetzt] an
ART	bestimmter oder unbestimmter Artikel	der, die, das, ein, eine
CARD	Kardinalzahl	zwei [Männer], [im Jahre] 1994
FM	Fremdsprachliches Material	[Er hat das mit ``] A big fish [`` übersetzt]
ITJ	Interjektion	mhm, ach, tja
KOUI	unterordnende Konjunktion mit ``zu" und Infinitiv	um [zu leben], anstatt [zu fragen]
KOUS	unterordnende Konjunktion mit Satz	weil, dass, damit, wenn, ob
KON	nebenordnende Konjunktion	und, oder, aber
KOKOM	Vergleichskonjunktion	als, wie
NN	normales Nomen	Tisch, Herr, [das] Reisen



News

Projektvorschlag zu Strukturierter Emotionsanalysis aus Text (SEAT) von der DFG genehmigt

15.09.2017

→ [Alle News sehen](#)

[Als Feed abonnieren](#)

Veranstaltungen

IMS student meeting -- November 8, 2017; lunch break; V 5.02

08.11.2017

Info meeting: exam/module registration -- November 16, 2017; lunch break; V 5.02

16.11.2017

IMS@unitag -- 22.11.2017

22.11.2017

→ [Alle Veranstaltungen sehen](#)

[Als Feed abonnieren](#)