

Korpusgestützte Grammatik

Datenbank attributive Adjektive

Der Datensatz attributive Adjektive 1.csv dient der Untersuchung zweier Variationsphänomene. Das erste Variationsphänomen ist das Schwanken zwischen schwacher und starker Flexion des zweiten Adjektivs in einer artikellosen Nominalphrase im Dativ Singular Maskulinum oder Neutrum: *mit hohem finanziellen Aufwand* vs. *mit hohem finanziellem Aufwand*. Das zweite Variationsphänomen betrifft die Reihenfolge der Adjektive untereinander (*bei trockenem[,] warmem Wetter* vs. *bei warmem[,] trockenem Wetter*).

Gesucht wurde in einem Ausschnitt von ca. 122 Mio. Tokens aus dem DEREKO die Abfolge

Präposition	Ø	erstes Adjektiv mit Endung -em	Ø	Ø	zweites Adjektiv mit Endung -em oder -en	Nomen
	Adverb		,	Adverb		

Der Datensatz enthält 1809 Sätze und nach dem Aussortieren von Fehlbelegen noch 1598 einschlägige Belege.

Wortgruppen, die diesem Muster entsprechen würden, sind etwa:

- *mit ziemlich langem, beinahe schwarzem/schwarzen Haar*
- *mit ziemlich langem schwarzem/schwarzen Haar*
- *mit langem, schwarzem/schwarzen Haar*
- *mit langem schwarzem/schwarzen Haar*

Die Tabellenspalten im Einzelnen:

Spalte	Überschrift	Beschreibung	Vorkommende Werte
A	Random	Zufallszahl (Reihenfolgenstudie: dient der Teilung des Datensatzes in zwei Hälften)	Bei einschlägigen Belegen (vgl. Spalte AU) natürliche Zahlen 1 bis 1598
B	Kandidat-ADJ	Angabe, welches der beiden Adjektive in der Reihenfolgenstudie als Kandidat zufällig gewählt wurde	1 (linkes Adjektiv), 2 (rechtes Adjektiv)

C	POS1	Die abhängige Variable in der Reihenfolgenstudie: Angabe, ob der Kandidat an erster Stelle (links) steht	0 (nein), 1 (ja)
D	SilbDiff.Kand-Konk	Silbenzahl (vgl. Spalten BE, BF) des Kandidaten minus Silbenzahl des Konkurrenten in der Reihenfolgenstudie	Ganze Zahlen -6 bis 8
E	CprobDiff.Kand-Konk	Bedingte Wahrscheinlichkeit (vgl. Spalten CF, CL), dass der Kandidat auftritt, minus bedingte Wahrscheinlichkeit, dass der Konkurrent auftritt, in der Reihenfolgenstudie	-0,59 bis 0,59
F	FreqDiff.Kand-Konk	DeReWo-Frequenzklasse (vgl. Spalten BJ, BK) des Kandidaten minus DeReWo-Frequenzklasse des Konkurrenten in der Reihenfolgenstudie	Ganze Zahlen -12 bis 12
G	Kand.Sema	Semantische Klasse (2-Klassen-System, vgl. Spalten AP, AQ) des Kandidaten in der Reihenfolgenstudie	int (intersektiv), non (nicht intersektiv)

H	Kand.Sem	Semantische Klasse (11-Klassen-System, vgl. Spalten AN, AO) des Kandidaten in der Reihenfolgenstudie	art-dem, art-indef, ordi, quali-farb, quali-rel, quali-sonst, ref-lok, ref-temp, rel-ber, rel-geo, rel-material (eine elfte Klasse, die nicht vorkam, wären Determinierer gewesen), u (nicht einzuordnen)
I	Kand.Part	Angabe, ob der Kandidat ein Partizip ist	lex (lexikalisiertes Partizip), nolex (nicht lexikalisiertes Partizip), nein (kein Partizip)
J	Kand.Adv	Angabe, ob der Kandidat durch ein Adverb erweitert ist (vgl. Spalten AF, AI)	0 (nein), 1 (ja)
K	Kand.Lemma	Lemma des Kandidaten (ersatzweise: Stamm)	<i>abgerundet bis zweit</i>
L	Kand.Cprob	Bedingte Wahrscheinlichkeit (vgl. Spalten CF, CL), dass der Kandidat auftritt	0 bis 0,60
M	Konk.Sema	Semantische Klasse (2-Klassen-System, vgl. Spalten AP, AQ) des Konkurrenten in der Reihenfolgenstudie	int (intersektiv), non (nicht intersektiv)
N	Konk.Sem	Semantische Klasse (11-Klassen-System, vgl. Spalten AN, AO) des Konkurrenten in der Reihenfolgenstudie	art-dem, art-indef, ordi, quali-farb, quali-rel, quali-sonst, ref-lok, ref-temp, rel-ber, rel-geo, rel-material (eine elfte Klasse, die nicht vorkam, wären Determinierer gewesen), u (nicht einzuordnen)
O	Konk.Part	Angabe, ob der Konkurrent ein Partizip ist	lex (lexikalisiertes Partizip), nolex (nicht lexikalisiertes Partizip), nein (kein Partizip)

P	Konk.Adv	Angabe, ob der Konkurrent durch ein Adverb erweitert ist (vgl. Spalten AF, AI)	0 (nein), 1 (ja)
Q	Konk.Lemma	Lemma des Konkurrenten (ersatzweise: Stamm)	<i>abgelebt bis zwölfjährig</i>
R	Konk.Cprob	Bedingte Wahrscheinlichkeit (vgl. Spalten CF, CL), dass der Konkurrent auftritt	0 bis 0,60
S	Index	Belegnummer	1 bis 1809
T	Zeitung	Sigle der Zeitung, aus der der Beleg stammt	A bis Z
U	Sigle	Textsigle	A00/APR.25746 bis Z98/811.06588
V	Quelle	Quellentyp	1 bis 12
W	Domain	Inhaltliche Domäne, die dem Text zugeordnet wurde	Fiktion, Kultur_Unterhaltung, Mensch_Natur, NA (= keine Angabe), Politik_Wirtschaft_Gesellschaft, Technik_Wissenschaft
X	Medium	Medium, aus dem der Text stammt	Buecher_Fachzeitschriften, Gesprochenes, Publikumspresse, Sonstiges
Y	Land	Land, aus dem der Text stammt	A (Österreich), CH (Schweiz), D (Deutschland), LU (Luxemburg)
Z	Region	Region, aus der der Text stammt	Mittelost, Mittelsued, Mittelwest, NA (keine Angabe), Nordost, Nordwest, Suedost, Suedwest, Ueberregional

AA	Register	Register, das dem Text zugeordnet wurde	Gebrauchstext, Literarisch, Presstext
AB	Jz	Jahrzehnt, in dem der Text erschienen ist	1990er, 2000er, 2010er, vor_1990
AC	Jahr	Jahr, in dem der Text erschienen ist	1942 bis 2014
AD	Vor	Satzkontext vor der Präpositionalphrase, die dem Suchmuster entspricht	Freitext
AE	APPR	Präposition	<i>an bis zwischen</i>
AF	ADV1	Adverb, das das erste Adjektiv modifiziert	0 (kein Adverb); <i>äußerst bis zunächst</i>
AG	ADJA1	Erstes attributives Adjektiv	<i>abgelebtem bis zwölfjährigem</i>
AH	Komma	Komma zwischen den beiden Adjektiven	0 (kein Komma), 1 (Komma)
AI	ADV2	Adverb, das das zweite Adjektiv modifiziert (oder Konjunktion, was eine Wertung als „nicht valide“ zur Folge hat)	0 (kein Adverb); <i>aber bis zusammen</i>
AJ	ADJA2	Zweites attributives Adjektiv	<i>aberwitzigen bis zwiespältigen</i>
AK	NN	Nomen (Kern der Nominalphrase)	<i>Abendessen bis Zwischenstopp</i>

AL	Nach	Satzkontext nach der Präpositionalphrase, die dem Suchmuster entspricht	Freitext
AM	Flexion	Starke/schwache Flexion des zweiten Adjektivs	m (stark), n (schwach)
AN	Sem1	Semantische Klasse (11-Klassen-System) des ersten Adjektivs	art-dem (Adjektiv, das einem Demonstrativum nahekommmt: <i>einige</i>), art-indef (Adjektiv, das einem Indefinitum nahekommmt: <i>besagt</i>), ordi (Ordinalzahladjektiv: <i>zweite</i>), quali-farb (Farbadjektiv: <i>rot</i>), quali-rel (ursprünglich qualitatives Adjektiv in relationaler Lesart, vgl. <i>stark</i> in <i>starke Raucherin</i>), quali-sonst (sonstiges qualifizierendes Adjektiv wie <i>groß</i>), ref-lok (referenzielles Adjektiv mit lokalem Bezug: <i>hiesig</i> , <i>dortig</i>), ref-temp (referenzielles Adjektiv mit temporalem Bezug: <i>gestrig</i> , <i>damalig</i>), rel-ber (Zugehörigkeitsadjektiv, das weder eine geografische Herkunft noch eine materielle Beschaffenheit bezeichnet: <i>finanziell</i>), rel-geo (Zugehörigkeitsadjektiv, das eine geografische Herkunft oder Zugehörigkeit bezeichnet: <i>sächsisch</i>), rel-material (Adjektiv, das eine materielle Beschaffenheit bezeichnet: <i>hölzern</i>); eine elfte Klasse, die nicht vorkam, wären Determinierer gewesen; u (nicht einzuordnen)
AO	Sem2	Semantische Klasse (11-Klassen-System) des zweiten Adjektivs	vgl. Spalte AN

AP	Sema1	Semantische Klasse (2-Klassen-System) des ersten Adjektivs	int (intersektiv, d. h. schnittmengenbildend wie <i>schweizerisch</i>), non (nicht intersektiv: subsektiv wie groß, privativ wie <i>ehemalig</i>)
AQ	Sema2	Semantische Klasse (2-Klassen-System) des zweiten Adjektivs	vgl. Spalte AP
AR	Und	„und“-Test-Ergebnis	0 (zwischen die beiden Adjektive lässt sich nicht ohne Bedeutungsveränderung die Konjunktion und einschieben), 1 (zwischen die beiden Adjektive lässt sich und einschieben), 2 (zwischen die beiden Adjektive lässt sich <i>und zwar, das heißt</i> oder <i>nämlich</i> einschieben)
AS	Tausch	Tauschtestergebnis	0 (die beiden Adjektive lassen sich nicht tauschen, ohne dass sich die Bedeutung ändert oder die Reihenfolge markiert klingt), 1 (die Adjektive lassen sich tauschen)
AT	Part2	Angabe, ob das zweite Adjektiv ein Partizip ist bzw. aus einem Partizip gebildet wurde; vgl. Spalte BO	nein (kein Partizip), lex (lexikalisiertes Partizip: <i>gewagt</i>), nolex (nicht lexikalisiertes Partizip: <i>gackernd</i>)
AU	valide	Angabe, ob der Beleg einschlägig ist	0 (Fehlbeleg), 1 (einschlägiger Beleg), CONJ (Beleg wurde nicht als einschlägig gewertet, weil zwischen den beiden Adjektiven eine Konjunktion steht)

AV	Anno_ADJA1	Automatische Annotation (POS: TreeTagger, morphosyntaktische Merkmale: Marmot) des ersten Adjektivs	Lemma/POS/morphosyntaktische Merkmale
AW	Stamm_A1	Stamm des ersten Adjektivs	<i>abgelebt</i> bis <i>zwölfjährig</i>
AX	Stamm_A2	Stamm des zweiten Adjektivs	<i>abgerundet</i> bis <i>zweckorientiert</i>
AY	Stamm_N	Stamm des Nomens	<i>Appeal</i> bis <i>Zoom</i>
AZ	Anno_ADV2	Automatische Annotation (POS: TreeTagger, morphosyntaktische Merkmale: Marmot) des Adverbs, das das zweite Adjektiv modifiziert	Lemma/POS/morphosyntaktische Merkmale
BA	Anno_ADJA2	Automatische Annotation (POS: TreeTagger, morphosyntaktische Merkmale: Marmot) des zweiten Adjektivs	Lemma/POS/morphosyntaktische Merkmale
BB	Anno_NN	Automatische Annotation (POS: TreeTagger, morphosyntaktische Merkmale: Marmot) des Nomens	Lemma/POS/morphosyntaktische Merkmale
BC	Silb1	Silbenzahl des ersten Adjektivs (ohne Flexionsendung) nach CELEX	1 bis 6

BD	Silb2	Silbenzahl des zweiten Adjektivs (ohne Flexionsendung) nach CELEX	1 bis 6
BE	Silb1.neu	Silbenzahl des ersten Adjektivs (inklusive Flexionsendung) nach einem Algorithmus, der bis auf einige Ausnahmen pro Vokal eine Silbe zählt	2 bis 8
BF	Silb2.neu	Silbenzahl des zweiten Adjektivs (inklusive Flexionsendung) nach einem Algorithmus, der bis auf einige Ausnahmen pro Vokal eine Silbe zählt	2 bis 10
BG	Differenz	Silb2 (Spalte BD) minus Silb1 (Spalte BC)	-4 bis 4; NA (keine Angabe)
BH	Differenz3	Werte aus Spalte BG, umcodiert auf 4 verschiedene Werte	positiv, negativ, 0, NA (keine Angabe)
BI	Differenz.neu	Silb2.neu (Spalte BF) minus Silb1.neu (Spalte BE)	-5 bis 8
BJ	Frequ1	DeReWo-Frequenzklasse des ersten Adjektivlemmas	6 (eher häufig) bis 20 (eher selten); NA (keine Angabe)

BK	Frequ2	DeReWo-Frequenzklasse des zweiten Adjektivlemmas	6 (eher häufig) bis 21 (eher selten); NA (keine Angabe)
BL	FrequNN	DeReWo-Frequenzklasse des Nomenlemmas	5 (eher häufig) bis 23 (eher selten)
BM	NN_m_Mitte	Anzahl von /m/ in der Mitte des Nomens (<mm> als ein /m/ gezählt, <m> oder <mm> am Ende des Nomens nicht gezählt)	0, 1, 2 (<i>Sommerhimmel</i>)
BN	NN_m_Ende	Angabe, ob das Nomen auf /m/ endet oder nicht	0 (nein), 1 (ja)
BO	Part 1	Angabe, ob das erste Adjektiv ein Partizip ist bzw. aus einem Partizip gebildet wurde; vgl. Spalte AT	nein (kein Partizip), lex (lexikalisiertes Partizip: <i>gewagt</i>), nolex (nicht lexikalisiertes Partizip: <i>gackernd</i>)
BP	Sem12	Kombination semantischer Klassen nach dem 11-Klassen-System (Verkettung des Wertes aus Spalte AN mit dem Wert aus Spalte AO)	art-demordi bis urel-geo
BQ	Sema12	Kombination semantischer Klassen nach dem 2-Klassen-System (Verkettung des Wertes aus Spalte AP mit dem Wert aus Spalte AQ)	intint (intersektiv + intersektiv), intnon (intersektiv + nicht intersektiv), nonint (nicht intersektiv + intersektiv), intnon (intersektiv + nicht intersektiv)

BR	Part12	Kombination von Wortbildungsarten nach dem Kriterium „Partizip oder nicht“ (Verkettung des Wertes aus Spalte BO mit dem Wert aus Spalte AT)	lexlex, lexnein, lexnolex, neinlex, neinnein, neinnolex, nolexlex, nolexnein, nolexnolex
BS	Sem1Num6	Wert des ersten Adjektivs bei der Einteilung in 6 Distanzklassen	1 (nahe am Nomen, tendenziell also rechts) bis 6 (entfernt vom Nomen, tendenziell also links): 1 Zugehörigkeitsadjektive (einschließlich solcher, die die geografische Herkunft bezeichnen); 2 Ordinalzahl- und Farbadjektive (wobei die Stellung der Ordinalzahladjektive so nahe am Nomen eine Besonderheit des vorliegenden Datensatzes ist), 3 relationale und sonstige qualifizierende Adjektive, 4 referenzielle Adjektive mit lokalem Bezug, 5 referenzielle Adjektive mit temporalem Bezug, 6 Adjektive, die Demonstrativa oder Indefinita nahekommen
BT	Sem2Num6	Wert des zweiten Adjektivs bei der Einteilung in 6 Distanzklassen	vgl. Spalte BS
BU	SemDist6	Sem1Num6 (Spalte BS) minus Sem2Num6 (Spalte BT)	-1 bis 5; NA (keine Angabe)
BV	SemDist	Werte von Spalte BU, umcodiert auf 4 verschiedene Werte	null, negativ, positiv, NA (keine Angabe)

BW	FreqDiff	Frequenzklasse des zweiten Adjektivs Frequ2 (Spalte BK) minus Frequenzklasse des ersten Adjektivs Frequ1 (Spalte BJ)	-8 bis 12; NA (keine Angabe)
BX	FrequDiff3	Werte aus Spalte BW, umcodiert in 4 verschiedene Werte	negativ, null, positiv, NA (keine Angabe)
BY	ADV	Angabe, ob das erste oder das zweite Adjektiv durch ein Adverb erweitert ist	ADV1, ADV2, (Leere)
BZ	Anfang_Fuellung	Angabe, ob Spalte AD gefüllt ist	0 (nein), 1 (ja)
CA	Genus	Genus des Nomens	n (Neutrum), m (Maskulinum), NA (keine Angabe)
CB	N.lemma.condprob	Nomenlemma (hilfsweise: Nomenstamm)	<i>Abendessen</i> bis <i>Zwischenstopp</i>
CC	adj1.lemma.condprob	Lemma (hilfsweise: Stamm) des ersten Adjektivs	<i>achtjährig</i> bis <i>zwölfjährig</i>
CD	f_ADJ1_N	Vorkommen des ersten Adjektivs als einziges attributives Adjektiv zum betreffenden Nomen im gesamten KoGra-Untersuchungskorpus (ca 7 Mrd. Tokens)	0 bis 183.093

CE	f_N	Vorkommen des Nomens im gesamten KoGra-Untersuchungskorpus (ca 7 Mrd. Tokens)	1 bis 16.722.239
CF	f_ADJ1 N	Bedingte Wahrscheinlichkeit, dass das erste Adjektiv auftritt: Wert aus Spalte CD geteilt durch Wert aus Spalte CE	0 bis 0,43
CG	f_ADJ1N_/N	Wert aus Spalte CF in anderem Zahlenformat	0 bis 0,43
CH	adj2.lemma.condprob	Lemma (hilfsweise: Stamm) des zweiten Adjektivs	<i>adlig bis zeitlich</i>
CI	f_ADJ2_N	Vorkommen des zweiten Adjektivs als einziges attributives Adjektiv zum betreffenden Nomen im gesamten KoGra-Untersuchungskorpus (ca 7 Mrd. Tokens)	0 bis 30.830
CJ	f_N	(identisch mit Spalte CE)	(identisch mit Spalte CE)
CK	f_ADJ2 N	Bedingte Wahrscheinlichkeit, dass das zweite Adjektiv auftritt: Wert aus Spalte CI geteilt durch Wert aus Spalte CE	0 bis 0,29

CL	f_ADJ2/N	Wert aus Spalte CK in anderem Zahlenformat	0 bis 0,29
CM	f_ADJ_N_diff	Wert aus Spalte CK (bedingte Wahrscheinlichkeit, dass das zweite Adjektiv auftritt) minus Wert aus Spalte CF (bedingte Wahrscheinlichkeit, dass das erste Adjektiv auftritt)	-0,42 bis 0,29
CN	Bedingte_Wahrscheinlichkeit_positiv	Angabe, ob der Wert aus Spalte CM positiv ist	FALSCH, WAHR

Autor(en)

Franziska Münzberg

Bearbeiter

Jana Mathy

Letzte Änderung

26. Okt. 2018

Link zum Artikel

<https://grammis.ids-mannheim.de/korpusgrammatik/6474>