TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

Einführung in die Angewandte Linguistik

Korpuslinguistik I

Prof. Dr. Simon Meier-Vieracker

simon.meier-vieracker@tu-dresden.de

Sitzungsverlauf

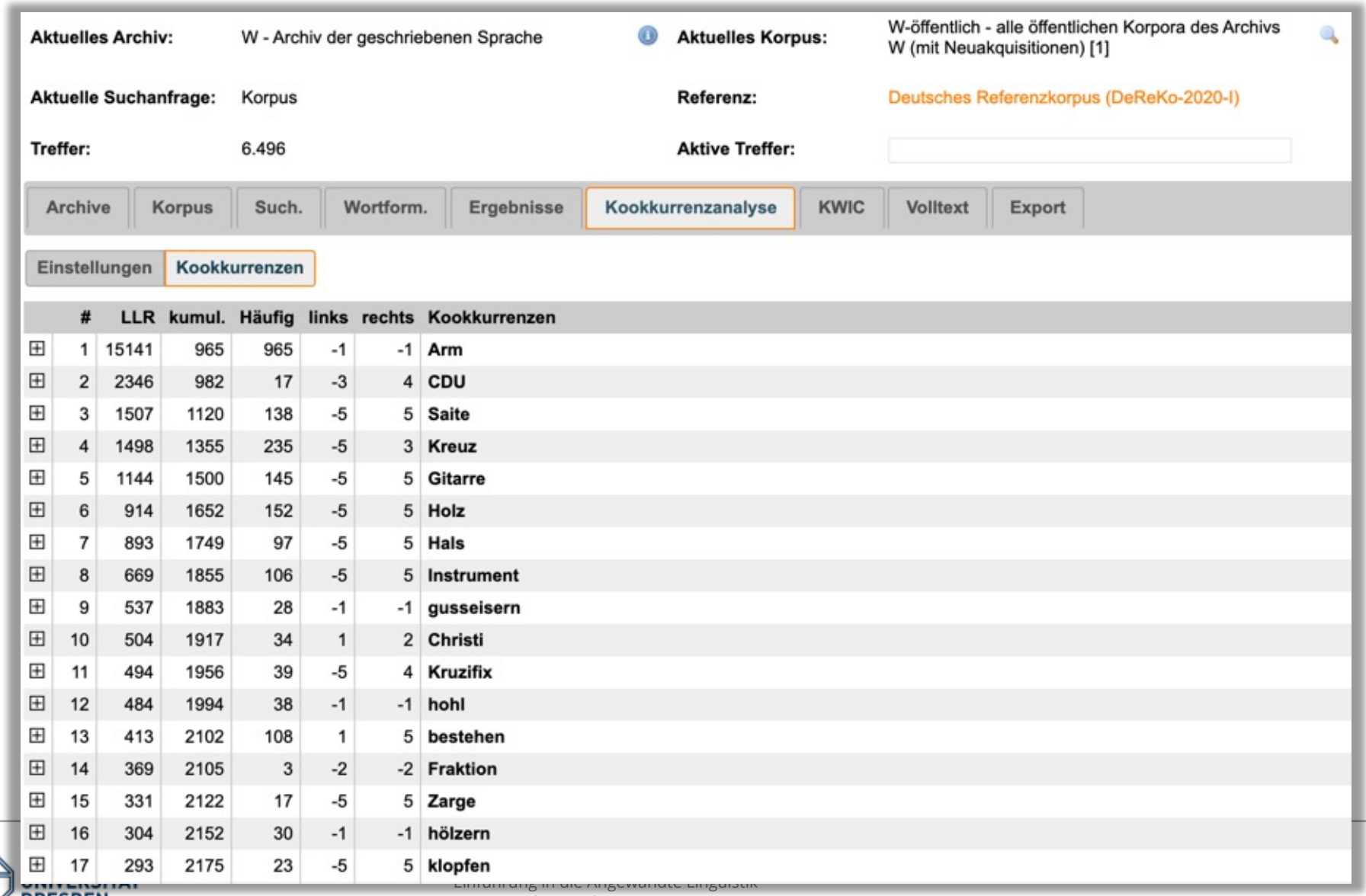
1. Was sind und wozu braucht man Korpora?
2. Referenz- und Spezialkorpora (nicht nur) des Deutschen
3. Korpora erstellen – aber wie?
4. Korpusaufbereitung: Annotationen und Metadaten

Lernziele

- Sie können erklären, was Korpora sind und wofür man sie in der Sprachwissenschaft braucht
- Sie können den Unterschied zwischen Referenz- und Spezialkorpora erklären.
- Sie können erklären, was Annotationen und Metadaten sind und wofür man sie braucht (und wie man es technisch lösen kann).
- Sie können den Unterschied zwischen Types und Tokens erklären.

Was ist ein Korpus?

Was ist ein Korpus?



Aktuelles Archiv: W - Archiv der geschriebenen Sprache
Aktuelles Korpus: W-öffentlich - alle öffentlichen Korpora des Archivs W (mit Neuakquisitionen) [1]
Aktuelle Suchanfrage: Korpus
Referenz: Deutsches Referenzkorpus (DeReKo-2020-I)
Treffer: 6.496
Aktive Treffer:

Archive | Korpus | Such. | Wortform. | Ergebnisse | Kookkurrenzanalyse | KWIC | Volltext | Export

Einstellungen | Kookkurrenzen

	#	LLR	kumul.	Häufig	links	rechts	Kookkurrenzen
⊞	1	15141	965	965	-1	-1	Arm
⊞	2	2346	982	17	-3	4	CDU
⊞	3	1507	1120	138	-5	5	Saite
⊞	4	1498	1355	235	-5	3	Kreuz
⊞	5	1144	1500	145	-5	5	Gitarre
⊞	6	914	1652	152	-5	5	Holz
⊞	7	893	1749	97	-5	5	Hals
⊞	8	669	1855	106	-5	5	Instrument
⊞	9	537	1883	28	-1	-1	gusseisern
⊞	10	504	1917	34	1	2	Christi
⊞	11	494	1956	39	-5	4	Kruzifix
⊞	12	484	1994	38	-1	-1	hohl
⊞	13	413	2102	108	1	5	bestehen
⊞	14	369	2105	3	-2	-2	Fraktion
⊞	15	331	2122	17	-5	5	Zarge
⊞	16	304	2152	30	-1	-1	hölzern
⊞	17	293	2175	23	-5	5	klopfen

Was ist ein Korpus?

- „Ein Korpus ist eine Sammlung schriftlicher oder gesprochener Äußerungen in einer oder mehreren Sprachen. Die Daten des Korpus sind digitalisiert, d.h. auf Rechnern gespeichert und maschinenlesbar. Die Bestandteile des Korpus, die Texte oder Äußerungsfolgen, bestehen aus den Daten selbst sowie möglicherweise Metadaten, die diese Daten beschreiben, und aus linguistischen Annotationen, die diesen Daten zugeordnet sind.“ (Lemnitzer/Zinsmeister 2015, 39)
- WICHTIG: Es heißt „**das** Korpus“ !!!!!!!!!!!!!!!!!!!!!!!!!!

Was ist ein Korpus?

- Drei Eigenschaften von linguistischen Korpora (Stefanowitsch 2020, 22f.)
  - the instances of language use contained in it are authentic;
  - the collection is representative of the language or language variety under investigation;
  - the collection is large.
- Korpuslinguistik als empirische Untersuchung von Sprachgebrauch auf der Basis von Korpora.
- „Corpus linguistics is the investigation of linguistic research questions that have been framed in terms of the conditional distribution of linguistic phenomena in a linguistic corpus.“ (Stefanowitsch 2020, 56)

Wozu Korpora?

- Der klassische Weg der Grammatikforschung operiert mit konstruierten Beispielsätzen und Akzeptabilitätsurteilen („grammatische Kompetenz“)

(1) Wir wissen, der Weg ist noch lang.

(2) *Wir wissen nicht, der Weg ist noch lang

- Aber: Wie verlässlich sind Akzeptabilitätsurteile, um etwas über Varietäten (abseits der prestigeträchtigen Standardvarietät) zu erfahren?

(3) Ich bin froh, wenn ich wieder zurück bin, **weil** Hamburg ist sehr vielfältig.

- Ohne Rückgriff auf Korpora würde *weil* mit Verbzweitstellung wohl als ‚nicht akzeptabel‘ beurteilt werden.

Wozu Korpora?

Treffer		
na	**mal**	nim **wird** des des signal die ganze zeit schwanke
uff masse isch in dem moment richtig	**weil**	es **isch** ja bloß de transistor überbrückt
zündfunke	**weil**	sunscht **geht** s kaputt
	weil	wir **werden** ja nach dieser theoretischen einheit dann
baba ob er noch was lesen mag	**weil**	ich **bin** jetzt echt müde
	weil	es **is** ja ich meine mh es is schon interessant
nachher man was weg genommen werden kann	**weil**	das **funktioniert** ja nach bestimmten regeln
das is wirklich ne interessante schlussfolgerung	**weil**	man **muss** glaub ich von zwei seiten rangehen ne man
meiner jacke ka ma das grün sehe	**weil**	rot **find** ich im moment nit okay danke
dann de er auf das ältere besteht	**weil**	er **sacht** halt das machema schon seit
	weil	sie **meinten** ähm
	weil	ich **mach** s ja im grund net
dir nichts äh oder freistellen au nich	**weil**	wir **brauchen** dich und
und s war eigentlich die eine woche	**weil**	wir **ham** des ja au zwei ma ta l ge
kämen sie dann noch klar	**weil**	oder dann **müssten** s eventuell die freizeitaktivitäte ein sch…
	weil	des **sin** se
	weil	wir **ham** s jetz
wenn ja eh die ganze leitunge ohmisch	**weil**	ich **hab** s ja als vorbedingung auch die leitunge ohmisch
	weil	wir **ham** ja hier noch ne rotierende hochspannungsverteilu…
	weil	die **bringt** so eine zarte störstrahlung rein wo jeder igeber

Wozu Korpora?

- In der Sprachgeschichtsforschung oder der Gesprächsanalyse, wo Daten nicht fingiert werden können, sind Korpora immer schon und notgedrungen die eigentliche Forschungsgrundlage.
- Auch die Lexikographie, besonders im Rahmen sog. Belegwörterbücher, stützt sich auf Korpora.
- Vergleichsweise neu sind dagegen die computergestützten Auswertungsmöglichkeiten und dementsprechend die Daten*mengen*, die nunmehr bewältigt werden können.

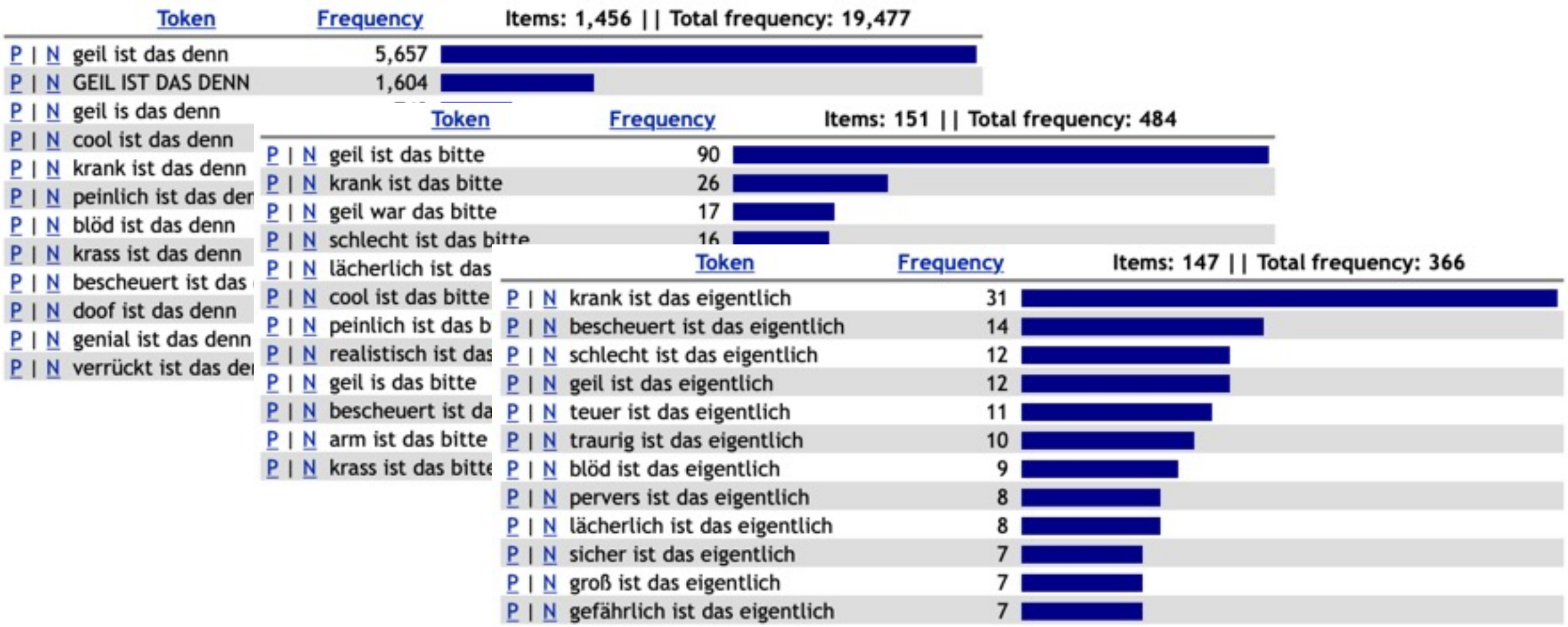
Wozu Korpora?

- Wie _____ ist das denn/bitte/eigentlich?

2009 " sarah biener meint : 3. März 2009 at 01:21	wie langweilig ist das denn	??? Wer will Boris noch sehen - diesen komischen
Freitag Morgen in der Wiederholung ansehen .	Wie bescheuert ist das denn	, angeblich kein Zimmer mehr frei , trotz eines
Deine Familie ist ja wirklich der Oberknaller !	Wie dreist ist das denn	? Haben die sich alle miteinander abgesprochen
LN 6 V Original von Handmade Original von LN 6 V	Wie geil ist das denn	... Alte Beiträge vom Handmade ... :tongue:
nachdenken ... der onkel hatte dreck am kragen .	wie peinlich ist das denn	? am 22. Januar 2008 um 9:03 pm Uhr Dito
von Schnauzer Rames und Frauchen Katja ...	wie toll ist das denn	, bitte ? Wir kenne uns ja eigentlich nur online
^^ daumen hoch ... tschaumitvau Says : sag mal	wie schlecht ist das denn	? was wirklich co ... tschaumitvau Says : und
Grüße Marianne :) SewCat 16.07.2010 , 06:12 ...	wie süß ist das denn	??? :) :) Ich habe mir früher immer sehnlichst
" ? :shok: FCBayern 20.11.2010 , 21:12	Wie affig ist das denn	? Wozu gibt es im Spiel denn bitte noch die "
ist toll ! * : lachboden 01.12.2008 , 17:32	Wie Geil ist das denn	: lach : lach Aber ich glaube bald , du
hin , dass ich es nach und nach editiere .	Wie dumm wäre das denn	, wenn ich schon alles vorher sage ? Da kann ich
schnellen flitzer-typen verprügelt ...	wie weit war das denn	? spielzeit müsste ungefähr 5 stunden gewesen
am arsch ist , vllt hält die den strom nicht mehr	wie alt ist die denn	? Marcel 86 30. Jul 2007 , 22:45 Also das selbe
scharf war wuste nicht mal das es der Zünder ist	wie geil ist das denn	!!! :ok: Gregor 18.10.2007 , 06:45 Lob : Hängst
Den fand ich auch nicht schlecht : uriblank : top	Wie geil ist das denn	? Da muss man erstmal drauf kommen . Gibts das
Irina sagt : 6. November 2009 um 10:50 Oh	wie geil ist das denn	??? Super ! Kompliment Das ist doch wirklich mal
ch kam . Meine Frau soll am 18 hin und ich am 19.	Wie bescheuert ist das denn	?? Unsere Arge ist 18 km von uns weg . Echt toll !
wiederbelebt werden . uriblank :rofl:	Wie geil ist das denn	? :lol: Passend dazu : Hotel for Dogs (xD)
Jess heult ? Es erklang nur ein leises Brummen .	Wie süß war das denn	, hatte er wegen mir geheult ? Ich sah zum
im Nehmen . Und jetzt erzähl doch mal ! Mensch ,	wie lang ist das denn	bereits her ?! Zehn Jahre ! Zehn volle Jahre ! Und

Quelle: https://www.webcorpora.org, DECOW16B (Austrian, German and Swiss German)

Wozu Korpora?



Auer, Peter (2016): „Wie geil ist das denn?“ In: Zeitschrift für germanistische Linguistik 44 (1), S. 69–92.

Referenz- und Spezialkorpora

- **Referenzkorpora** versuchen, eine Sprache möglichst repräsentativ zu dokumentieren
- Problem: Nur bei toten Sprachen, und auch hier nur in der Theorie, kann wirklich jede (schriftliche) Äußerung erfasst werden.
- **Spezialkorpora** versuchen dagegen, einen bestimmten, möglicherweise eng umgrenzten Sprachausschnitt zu erfassen…
- …allerdings so, dass sie *für diesen (im Untersuchungsziel entsprechend zu begründenden) Sprachausschnitt* repräsentativ sind.

Referenz- und Spezialkorpora

- Deutsches Referenzkorpus (DeReKo) des Mannheimer Instituts für Deutsche Sprache:
  - 46,9 Milliarden laufende Wörter (ein durchschnittlicher Roman von 200 Seiten hat 80.000–100.000 Wörter)
  - Verschiedenste Textsorten und Kommunikationsbereiche, von Grimms Märchen über Plenarprotokolle bis hin zu Wikipedia-Diskussionen, vor allem aber Pressetexte.
  - Über die Rechercheplattform COSMAS II sind komplexe Korpusabfragen einschließlich statistischer Auswertungen möglich.
  - http://www.ids-mannheim.de/cosmas2/web-app/

Referenz- und Spezialkorpora

- Digitales Wörterbuch der deutschen Sprache DWDS (www.dwds.de)
  - 27 Milliarden laufende Wörter (lemmatisiert und POS-annotiert, s.u.)
  - Besonderheit DWDS-Kernkorpus: Zeitlich und nach Textsorten (einigermaßen) ausgewogenes Korpus mit Texten aus dem gesamten 20. Jahrhundert
  - Über diese Adresse u.a. auch zugänglich: Das deutsche Textarchiv (DTA) mit 157 Mio. lfd. Wörtern aus dem Zeitraum 1600–1900.
  - Flexible Abfragesyntax für den Zugriff auf die annotierten Daten.

Referenz- und Spezialkorpora

- Beispiel: "weil $p=ADV $p=VVFIN“ im Blog-Korpus
- *weil* gefolgt von einem Adverb gefolgt von einem finiten Verb

1:	2005	2005___10_	… runtermachen , kann ich das mit dem itunes machen ??	**weil**	**da gibts** ja die taste wiederherstellen oder so. geht das ?
2:	2005	2005___10_	…solut keinen Plan, wo ich das Design umschreiben muss,	**weil**	**eigentlich stimmt** meiner Ansicht nach soweit alles.
3:	2005	2005___11_		**Weil**	**da kommt** an den Umkleiden stehen irgendwie noch blö.
4:	2005	2005___11_		**Weil**	**da kommt** an den Umkleiden stehen irgendwie noch blö.
5:	2006	2006___03_	Schwerwiegende Fragen,	**weil**	**eigentlich tuts** der Rechner für mich, die aktuelle Kühlu.
6:	2006	2006___04_	…n, ich sass lediglich dahinter, am Fenster hinten rechts,	**weil**	**da gibt** es Internet und somit kann ich die Zeit wunderb..
7:	2006	2006___08_	Besser ist auch immer man hat den Betrag passend,	**weil**	**sonst stimmt** es vielleicht nicht.
8:	2006	2006___08_	Ware muss man auch noch nach prüfen,	**weil**	**sonst fehlt** was.
9:	2007	2007___02_		**Weil**	**manchmal kommt** man lange nicht auf die jeweilige For
10:	2007	www____20(		**Weil**	**dann fällt** es ja ganz leicht runter.
11:	2007	2007___05_	Wollte aber jetzt dennoch mal kommentieren,	**weil**	**sonst schreibt** man so ins Leere hinein.
12:	2007	martha_arge	(	**Weil**	**oben steht** 04. Juni, heute ist aber erst der 02. – rätselh
13:	2007	2007___08_	Sry habe ich vergessen ist bicheon so wichtig für nuker	**weil**	**sonst mache** ich mir wenn es nich wichtig wer 90fire ice
14:	2007	2007___08_	…og meint einen eigenen Werbe Spot drehen zu müssen,	**weil**	**irgendwann merkt** das keiner mehr.
15:	2007	2007___12_	…chon gewundert, dass wir hier Bilder davon sehen dürfen,	**weil**	**sonst gilt** doch in Museen “Fotografieren verboten”.
16:	2007	2007___12_	…chon gewundert, dass wir hier Bilder davon sehen dürfen,	**weil**	**sonst gilt** doch in Museen “Fotografieren verboten”.
17:	2008	2008___02_		**Weil**	**dann entfallen** euch diese dummen Fragen !!! o_0

Referenz- und Spezialkorpora

- TenTenCorpora via SketchEngine (www.sketchengine.eu)
  - Automatisiert erhobene Korpora mit Webtexten verschiedener Sprachen
  - deTenTen13: 19 Milliarden Tokens
  - Über Ihren Uni-Account frei zugänglich, inkl. Nutzung der sehr umfangreichen statistischen Auswertungsmöglichkeiten
  - Auch Upload eigener Korpora ist möglich!

Referenz- und Spezialkorpora

- British National Corpus (http://www.natcorp.ox.ac.uk):
  100 Mio. Wörter in geschriebener und gesprochener Sprache, zeitlich und textsortenbezogen ausgewogen
- Korpora gesprochener Sprache: Datenbank Gesprochenes Deutsch (https://dgd.ids-mannheim.de): 4748 Transkripte zu 3245 Stunden Audio- und Videomaterial, darunter Unterrichtskommunikation, Prüfungsgespräche, Tischgespräche, Schlichtungsgespräche usw. usf.

Referenz- und Spezialkorpora

- Einige Spezialkorpora für die deutsche Sprache:
  - Dortmunder Chat-Korpus (http://www.chatkorpus.tu-dortmund.de) mit 1,6 Mio. lfd. Wörtern
  - Mobile Communication Database MoCoDa2 (https://db.mocoda2.de/c/home) mit 532 WhatsApp-Chats mit 33.512 Nachrichten
  - Text+Berg-Korpus mit den Jahrbüchern des Schweizer Alpenvereins 1864–2014 mit 53 Mio. lfd. Wörtern (https://textberg.ch/site/de/korpora/)
  - Multilinguale Korpora zur Fußballlinguistik (https://fussballlinguistik.de/korpora/) mit Livetickern, Spielberichten, Taktikanalysen usw., 44 Mio. lfd. Wörter


BREAKFAST – KEBABS
HAMBURGER BAR
SANDWICHES
COFFEE
HOUSE
Küche vs. Restaurant
Essen gehen oder selber kochen?

Essen gehen oder selber kochen?

- Wer essen geht, spart Zeit und Mühe und kann Dinge probieren, die er/sie selbst nicht oder nicht in dieser Qualität zubereiten kann.
- Wer selbst kocht, weiß genau was drin ist, kann auf jeden Spezialwunsch eingehen und hat dieses unvergleichliche Selfmadegefühl.
- Wer auf bestehende Korpora zurückgreift…
- Wer Korpora selber erstellt…

Korpora erstellen, aber wie?

- Gefahr des Zirkelschlusses bei rein linguistischen Auswahlkriterien.
- Bsp: Sammlung von Texten, die *weil* mit Verbzweitstellung enthalten, um *weil* mit Verbzweitstellung zu untersuchen??

→ Korpora müssen immer *auch* quantifizierende Aussagen zum untersuchten Phänomen ermöglichen.

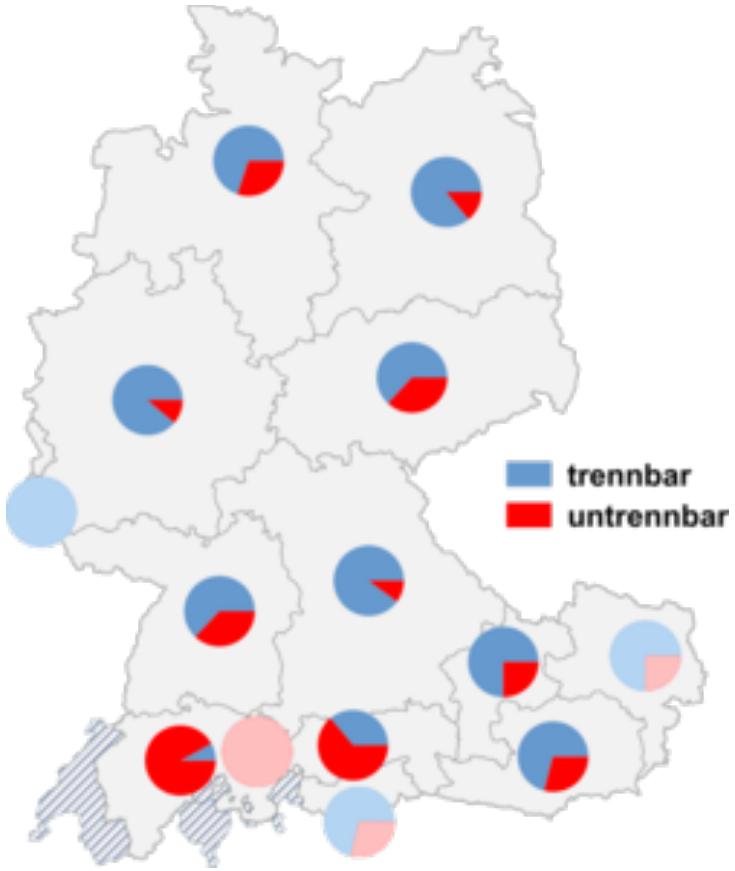
- Bsp. satzinitiales *also* im Vorvorfeld ("also WITH $.=0 $p=PPER")
  „Also ich bevorzuge die klassische Variante…“

Die Zeit (1946–2018)	Mode- und Beautyblogs (1997–2020)
267 Treffer	16364 Treffer
0.47 pro Mio. Wörtern	52.7 pro Mio. Wörtern

Korpora erstellen, aber wie?

- Alternative: Pragmatische Kriterien wie z.B.
  - Domäne (öffentlich, privat, akademisch usw.)
  - Medialität (mündlich, schriftlich) und Adressierung (monologisch, interaktiv)
  - Textsorten mit bestimmten Textfunktionen (Werbeanzeigen, Geschäftsbriefe, Fußballliveticker usw.)
- Auch mögliche Auswahlkriterien: Entstehungsort und Entstehungsdatum
- Ermöglicht die Untersuchung „in terms of the conditional distribution of linguistic phenomena in a corpus“ (Stefanowitsch 2020, 56)

Nochmal: anerkennen



MORPH(PRON per irr) /w0 <sa> &anerkennen /w0 MORPH(VRB fin v)

Sie anerkennt aber, dass sie auch in der FDP «nicht nur Freund
Er anerkennt die Leistung der Familie Berger für den Tourismus

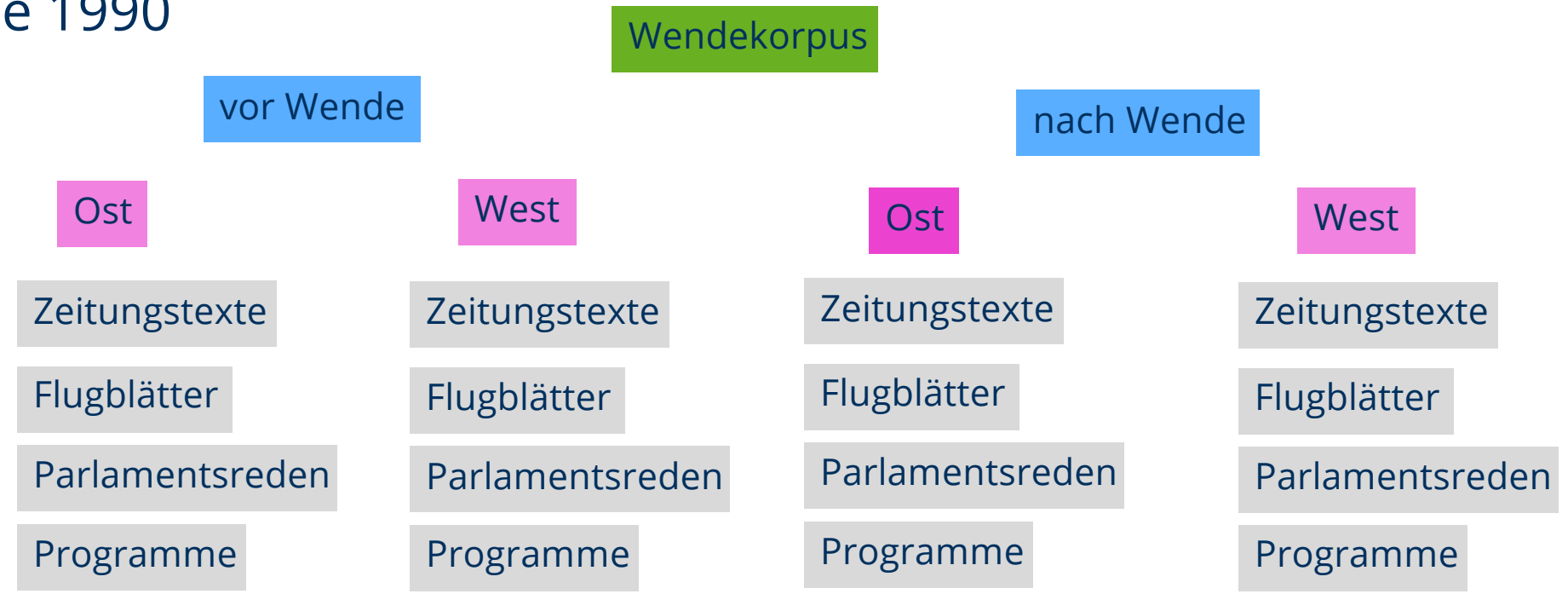
	Treffer	rel. Häuf.	Texte	von	bis	Land
⊞	4	0.0303 pMW	4	2012	2013	A
⊞	99	0.5806 pMW	99	2010	2014	CH
⊞	11	0.0149 pMW	11	2010	2014	D
	114	**0.1095 pMW**	**114**	**2010**	**2014**	**3 Länder**

	Treffer	rel. Häuf.	Texte	von	bis	Quelle
⊞	3	0.0348 pMW	3	2010	2013	Braunschweiger Zeitung
⊞	63	0.6034 pMW	63	2010	2013	St. Galler Tagblatt
⊞	3	0.0279 pMW	3	2010	2013	Süddeutsche Zeitung
	114	**0.1095 pMW**	**114**	**2010**	**2014**	**9 Quellen**

http://mediawiki.ids-mannheim.de/VarGra/index.php

Korpora erstellen, aber wie?

- Korpora können intern gegliedert werden, um kontrastive Fragestellungen zu bearbeiten.
- Beispiel Wendekorpus: 3387 Texte aus Ost und West von Mitte 1989 bis Ende 1990



Korpora erstellen, aber wie?

- Korpora sind immer Stichproben, sollen aber repräsentativ sein, also Schlüsse auf die Grundgesamtheit zulassen.
- Problem: Wenn man die Grundgesamtheit nicht kennt, kann ein noch so umfangreiches Korpus nicht im eigentlichen Sinne repräsentativ sein.
- Darum sind Googlesuchtrefferzahlen *keine* validen Belege für die Häufigkeit eines Ausdrucks!!
- *Die* ideale Korpusgröße gibt es nicht, auch sehr kleine Korpora können je nach Fragestellung und Auswertungsmethode interessant und ausreichend sein.

Annotation

- Durch Annotation werden den sprachlichen Daten interpretative linguistische Informationen hinzugefügt.
- Typische Schritte: Tokenisierung, part-of-speech-Tagging und Lemmatisierung.
- Dies übernehmen für gewöhnlich Programme, die etwa beim pos-Tagging mit bis zu 98%iger Wahrscheinlichkeit zuverlässig sind.
- Durch Annotationen werden Korpora vielseitiger einsetzbar und Suchen im Text können abstrakter sein (z.B. Suche nach allen Wortformen des Verbs *sein* oder nach allen Adjektiven).

Annotation

- Tokenisierung: Aufsplittung der laufenden Kette sprachlicher Zeichen in die einzelnen Tokens.

```
Aufsplittung
der
laufenden
Kette
sprachlicher
Zeichen
in
die
einzelnen
Tokens
.
```

Annotation

- part-of-speech-Tagging: Jedem Wort werden (automatisiert) die Wortarten nach einem standardmäßigen Tagset (hier STTS) zugewiesen.

```
Aufsplittung        NN
der                 ART
laufenden           ADJA
Kette               NN
sprachlicher        ADJA
Zeichen             NN
in                  APPR
die                 ART
einzelnen           ADJA
Tokens              NN
.                   $.
```

Annotation

- Lemmatisierung: Jedem Wort wird die unflektierte Grundform zugewiesen.

```
Aufsplittung    NN      Aufsplittung
der             ART     die
laufenden       ADJA    laufend
Kette           NN      Kette
sprachlicher    ADJA    sprachlich
Zeichen         NN      Zeichen
in              APPR    in
die             ART     die
einzelnen       ADJA    einzeln
Tokens          NN      Token
.               $.      .
```

Part-of-Speech (XPOS):

	PPER	VVFIN	PRF	$,	KOUS	PPER	ADV	APPR	NE	APPR	NN	NE	ADV	ART	NN	VAFIN	$.
1	Ich	freue	mich	,	dass	wir	hier	in	Dresden	mit	Frau	Staudinger	jetzt	eine	Rektorin	haben	.

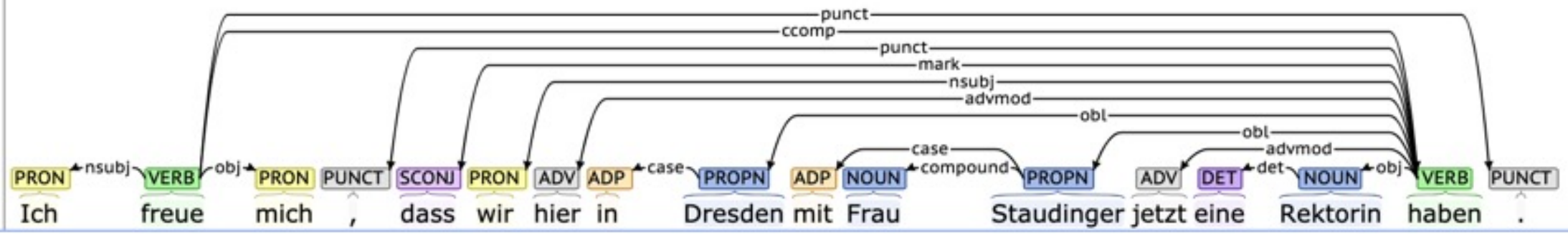
Lemmas:

	ich	freuen	ich	,	dass	wir	hier	in	Dresden	mit	Frau	Staudinger	jetzt	ein	Rektorin	haben	.
1	Ich	freue	mich	,	dass	wir	hier	in	Dresden	mit	Frau	Staudinger	jetzt	eine	Rektorin	haben	.

Named Entity Recognition:

		LOC							PER								
1	Ich	freue	mich	,	dass	wir	hier	in	Dresden	mit	Frau	Staudinger	jetzt	eine	Rektorin	haben	.

Universal Dependencies:



Ausprobieren unter: http://stanza.run/

Types und Tokens

- Bsp.: Der Koalitionsvertrag 2018 ist 69.128 Tokens lang (69128 Zeilen nach der Tokenisierung)…
- Aber nur 11.019 verschiedene Wörter → Types
- Häufigste Types:
  . und die , der wir für in werden wollen den

(1) Und ich versprach ihr, dass ich ihr das Buch wiedergeben würde.

→ 13 Tokens (mit Satzzeichen), aber nur 11 Types.

Metadaten

- Textinformationen, die die Beschaffenheit des Textes gemäß der Kriterien der Korpuszusammenstellung beschreiben, aber nicht Teil des Textes selbst sind.
- Bsp: Textsorte, Publikationsdatum, Quelle, Medialität, Geschlecht des/der Autor*in, …

Ein bisserl Technik

- Aus Nutzendenperspektive sind Metadaten oft Teil des Textes:

SACHSEN

06.05.2019 15:16 Uhr

War das Mathe-Abitur zu schwer?

Sächsische Schüler hatten Probleme mit Aufgaben in Geometrie und Stochastik. Auch in anderen Ländern gibt es Beschwerden.

von Andrea Schawe

Dresden. Sächsische Schüler aus dem Grund- und Leistungskurs beschweren sich über zu hohe Anforderungen bei den Abiturprüfungen in Mathematik. Etwa 10.800 Schüler waren zur Prüfung am vergangenen Freitag zugelassen. Der Landesschülerrat sprach von „einigen Beschwerden“ über Aufgaben im Bereich Geometrie und Stochastik und von zeitlichen Engpässen.

Ein bisserl Technik: HTML

- In digitalen Texten sind aber (im Hintergrund) reiner Text und Metadaten mit Informationen über den Text getrennt (Lobin 2014, 88):

```
<p class="article-section">Sachsen</p>
<p class="article-published-at">06.05.2019 15:16 Uhr</p>
<h1>War das Mathe-Abitur zu schwer?</h1>
<div class="article-lead"><p>Sächsische Schüler hatten Probleme mit Aufgaben in
Geometrie und Stochastik. Auch in anderen Ländern gibt es Beschwerden.</p></div>
<div class="article-author clearfix">Andrea Schawe</div>
<div class="article-content">
<p><b>Dresden.</b> Sächsische Schüler aus dem Grund- und Leistungskurs beschweren sich
über zu hohe Anforderungen bei den Abiturprüfungen in Mathematik. Etwa 10.800 Schüler
waren zur Prüfung am vergangenen Freitag zugelassen. Der Landesschülerrat sprach von
„einigen Beschwerden“ über Aufgaben im Bereich Geometrie und Stochastik und von
zeitlichen Engpässen.<br></p>
```

Ein bisserl Technik: XML

```
<text sitename="Sächsische Zeitung" title="War das Mathe-Abitur
 zu schwer?" author="Andrea Schawe" date="2019-05-06"
source="https://www.saechsische.de/plus/war-das-mathe-abitur-zu-schwer-
5067305.html">
<h1>
War             VAFIN  sein
das             ART    die
Mathe-Abitur    NN     Mathe-Abitur
zu              PTKA   zu
schwer          ADJD   schwer
?               $.     ?
</h1>
<p>
Sächsische      ADJA   sächsisch
Schüler         NN     Schüler
hatten          VAFIN  haben
Probleme        NN     Problem
mit             APPR   mit
```

Metadaten

Annotationen

Primärdaten

Lernen Sie programmieren!

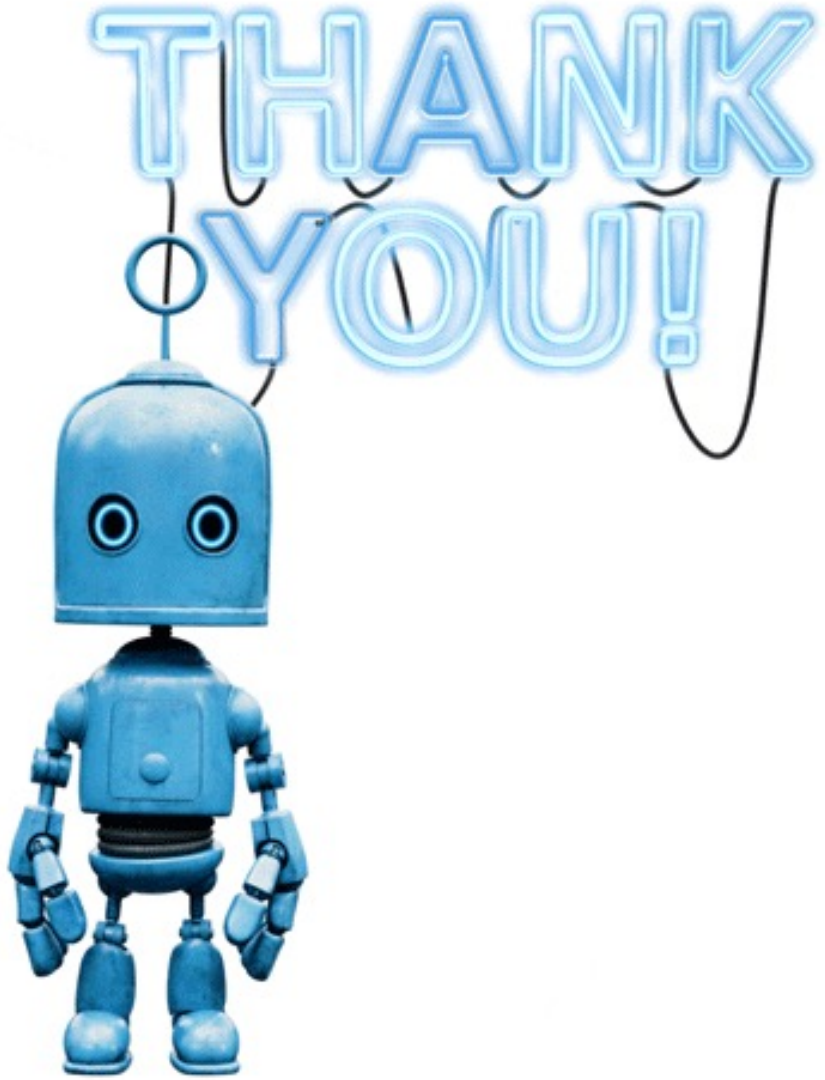
Zitierte Literatur

Auer, Peter (2016): „Wie geil ist das denn?“ In: Zeitschrift für germanistische Linguistik 44 (1), S. 69–92.

Lemnitzer, Lothar/Zinsmeister, Heike (2015): Korpuslinguistik. Eine Einführung. 3. Aufl. Tübingen: Narr Francke Attempto. (= Narr Studienbücher).

Lobin, Henning (2014): Engelbarts Traum: Wie der Computer uns Lesen und Schreiben abnimmt. Frankfurt: Campus Verlag.

Stefanowitsch, Anatol (2020): Corpus linguistics: A guide to the methodology. Berlin: Language Science Press. (= Textbooks in Language Sciences).


THANK
YOU!