

Gefördert durch



# Gingko (Geschriebenes ingenieurwissenschaftliches Korpus) Handbuch

Annette Portmann, Caren Brendel, Anna Sakhno, Mengxin Wu

Version 1.0

Stand: 01. Juli 2022

<https://www.uni-leipzig.de/gingko>

Bitte zitieren Sie das Handbuch wie folgt: Portmann, A., Brendel, C., Sakhno, A. & Wu, M. (2022).  
Gingko (Geschriebenes ingenieurwissenschaftliches Korpus): Handbuch.

## Korpusteam

Christian Fandrych (*Projektleitung*)

Antje Heine (*Projektleitung*)

Annette Portmann

Marlene Rummel

Lars Schirrmeister

Nina Suppus

Sarah Brauer

Caren Brendel

Paul Knötgen

Bárbara Mendoza Sánchez

Anna Sakhno

Mengxin Wu

# Inhalt

Abbildungsverzeichnis.....	1
Tabellenverzeichnis.....	1
1 Einleitung .....	2
2 Daten: Texte, Metadaten, Annotationen .....	3
2.1 Texte.....	3
2.2 Metadaten.....	4
2.2.1 Übersicht über die Metadatenvariablen.....	5
2.2.2 Informationen zu journalistischen Textsorten.....	7
2.3 Annotationen.....	8
2.3.1 Linguistische Annotationen .....	8
2.3.2 Annotationen zur Dokumentstruktur .....	12
3 Korpusnutzung in KorAP .....	12
3.1 Anmeldung für die IDS-Dienste KorAP und COSMAS II .....	12
3.2 Kennenlernen der Funktionen von KorAP.....	13
3.3 Korpusassistent: Eingrenzen der Textauswahl, Korpusstatistik, Suchanfragesprache ...	13
3.4 Stellen einer Suchanfrage.....	15
3.5 Trefferansicht.....	16
3.6 Abfragetricks – Zusammenstellen einer Suchanfrage über Trefferansicht .....	17
3.7 Datenexport.....	19
4 Zitieren des Korpus .....	19
Literaturverzeichnis.....	<b>Fehler! Textmarke nicht definiert.</b>
Anhang: Beispielsuchanfragen für Gingko in KorAP .....	21
Substantiv.....	21
Verb .....	22
Adjektiv .....	24
Präposition.....	25
Konjunktion .....	26

## Abbildungsverzeichnis

Abbildung 1: Belegstelle mit Metadaten und Annotationen in KorAP.....	3
Abbildung 2: Metadaten in Gingko .....	5
Abbildung 3: Annotationen eines Tokens .....	9
Abbildung 4: Satzannotation .....	10
Abbildung 5: Annotation von Satzkonstituenten.....	10
Abbildung 6: Annotation von Abhängigkeitsrelationen.....	10
Abbildung 7: KorAP-Tour und Einführung.....	13
Abbildung 8: Auswahl einer Metadatenvariable .....	14
Abbildung 9: Auswahl von Gingko über „corpusSigle“ .....	14
Abbildung 10: Auswahl von Gingko über „gingkoCollectionShort“ .....	14
Abbildung 11: Korpusstatistik.....	15
Abbildung 12: Auswahl der Suchanfragesprache .....	15
Abbildung 13: Übernehmen eines Suchausdrucks aus der KorAP-Einführung.....	16
Abbildung 14: Trefferanzeige ohne Glimpse-Funktion .....	16
Abbildung 15: Metadaten und Annotationen eines Suchtreffers .....	17
Abbildung 16: Neue Suchbedingung über Auswahl der Metadaten.....	18
Abbildung 17: Neue Suchanfrage über Auswahl der Tokenannotationen .....	18
Abbildung 18: Ergebnisexport .....	19

## Tabellenverzeichnis

Tabelle 1: Metadatenvariablen und beispielhafte Werte .....	7
Tabelle 2: Übersicht journalistischer Textsorten .....	8
Tabelle 3: Annotationen zur Dokumentstruktur.....	12

# 1 Einleitung

Gingko ist ein Korpus der geschriebenen ingenieurwissenschaftlichen Fachsprache (Schirrmeister et al. 2021). Es beinhaltet wissenschaftliche Artikel der Zeitschriften *Automobiltechnische Zeitschrift* (ATZ) und *Motortechnische Zeitschrift* (MTZ) des Springer-Verlags, zwei renommierten deutschsprachigen Fachzeitschriften im Bereich der Automobiltechnik. Das Korpus umfasst 2498 Texte der Jahrgänge 2007 bis 2016; dies entspricht 4.667.656 Millionen Token. Als digitale Textsammlung bietet das Korpus Forschenden eine umfangreiche Datengrundlage, die neben qualitativen Arbeiten auch die Beantwortung quantitativer Fragestellungen über die (Fach-)Wissenschaftssprache ermöglicht. Für Studierende der Ingenieurwissenschaften und für Lehrende in studienvorbereitenden oder -begleitenden Kursen für Deutsch als fremde Wissenschaftssprache stellt das Korpus eine Ressource für Beispiele von Wörtern und Strukturen im Kontext authentischer Texte dar. Gingko ist im *Deutschen Referenzkorpus* (DeReKo; Kupietz et al., 2018) am *Leibniz-Institut für Deutsche Sprache* (IDS) veröffentlicht und über die Korpustools *KorAP* (Kupietz et al., 2020) und COSMAS II kostenlos durchsuchbar.

Als vergleichbare, freizugängliche Korpora für die Wissenschaftssprache Deutsch können die als wissenschaftlich gekennzeichneten Sukorpora der DWDS-Kernkorpora (v. a. Lexikonartikel und geisteswissenschaftliche Texte; Berlin-Brandenburgische Akademie der Wissenschaften, o. J.), sowie die DWDS-Webkorpora zu Jura und Medizin (fachspezifische, populärwissenschaftliche Texte), das *Fachsprachenkorpus* und *spektrumdirekt* aus DeReKo (Artikel einer populärwissenschaftlichen Online-Zeitschrift) genannt werden.

Das Korpus wurde im Rahmen eines DFG-geförderten Projekts mit dem Kennzeichen HE 7157/1-1 erstellt, zunächst an der Universität Greifswald (2017–2019, Leitung: Jun.-Prof. Antje Heine) unter dem Titel „Muster in der Sprache der Ingenieurwissenschaften. Eine korpusbasierte konstruktionsgrammatische Analyse“, später fortgeführt am Herder-Institut der Universität Leipzig (2020–2022, Leitung: Prof. Christian Fandrych) unter der Bezeichnung „Muster in der Sprache der Ingenieurwissenschaften – Gingko (Geschriebenes ingenieurwissenschaftliches Korpus“, [www.uni-leipzig.de/gingko](http://www.uni-leipzig.de/gingko)). Dabei wurde mit dem *Karlsruher Institut für Technologie* und dem IDS zusammengearbeitet. Im Projekt wurden anhand von Gingko unpersönliche Ausdrucksmuster (Heine et al., 2019), figurative Verben (Heine, 2019), semantische Relationen (Heine & Rummel, 2019; Rummel & Heine, 2019) und Attributrealisierungen (Meißner, 2022) untersucht.

Im vorliegenden Handbuch werden in Kapitel 2 zunächst die Texte, Metadaten und Annotationen in Gingko beschrieben. Kapitel 3 erläutert die Nutzung des Korpus in KorAP. Kapitel 4 nennt Empfehlungen zum Zitieren des Korpus. Ergänzend zum Handbuch geben auch die Videotutorials auf [www.uni-leipzig.de/gingko](http://www.uni-leipzig.de/gingko) und die Beispielsuchanfragen im Anhang (S. 21) einen Überblick über den Zugriff auf Gingko.

## 2 Daten: Texte, Metadaten, Annotationen

Die Daten des Korpus bestehen aus Texten, Metadaten und Annotationen. Metadaten sind die Informationen über die Texte, z. B. die Namen der Autor:innen etc. Annotationen stellen Interpretationen der Texte in Form linguistischer Kategorien, wie Wortart oder Lemma, dar (Hirschmann, 2019, S. 2). Sie beziehen sich in Gingko auf Wörter, Phrasen und Sätze. Texte, Metadaten und Annotationen sind in der Trefferansicht in KorAP abrufbar (Kapitel 3.5; Abbildung 1).

diese bei konventioneller Motorisierung und erst recht bei Hybridantrieben völlig genügen? Fischer: " Mir kann kein Mensch erzählen, dass eine Vielgang-Automatik keine Mehrkosten verursacht. " Bei zunehmender Hybridisierung werden sich zudem Wertigkeiten in der Betrachtung der Getriebeeigenschaften verschieben. Die E-Maschine kann als <b>Anfahrelement</b> dienen, die Konkurrenz zwischen Drehmomentwandler und Kupplungen verliert somit an Bedeutung. Die elektromotorische Unterstützung wird darüber hinaus die subjektive Wahrnehmung von Gangsprüngen günstig beeinflussen. Der Wirkungsgrad als ein entscheidendes Kriterium könnte stärker als bisher in den Vordergrund treten. Auf der																											
<b>author</b>	Goppelt, Gernot	<b>availability</b>	QAO-NC	<b>corpusEditor</b>	Prof. Dr. Christian Fandrych, Leipzig University; Jun.-Prof. Dr. Antje Heine	<b>corpusExter...</b>	Gingko-Webseite an der Universität Leipzig	<b>corpusInter...</b>	IDS webpage on Gingko in the DeReKo archive	<b>corpusSigle</b>	ATZ14																
<b>corpusTitle</b>	Gingko - Geschriebenes Ingenieurwissenschaftlic...	<b>distributor</b>	Institut für Deutsche Sprache	<b>foundries</b>	<ul style="list-style-type: none"> <li>corenlp</li> <li>corenlp/constituency</li> <li>corenlp/morpho</li> <li>corenlp/sentences</li> </ul>	<b>gingkoArtic...</b>	doi:10.1007/s35148-014-0033-y	<b>docSigle</b>	ATZ14/FEB	<b>docTitle</b>	ATZ - Automobiltechnische Zeitschrift, Februar 2014																
<b>gingkoColle...</b>	Gingko - Geschriebenes Ingenieurwissenschaftliches Korpus	<b>gingkoColle...</b>	Gingko	<b>gingkoGenr...</b>	journalistisch	<b>gingkoGenr...</b>	Tagungsbericht	<b>gingkoLem...</b>	yes	<b>gingkoSource</b>	ATZ - Automobiltechnische Zeitschrift																
<b>indexCreati...</b>	2022-03-09	<b>indexLastM...</b>	2022-03-09	<b>language</b>	de	<b>pubDate</b>	2014	<b>pubPlace</b>	Wiesbaden	<b>pubPlaceKey</b>	DE																
<b>srcPages</b>	8-9	<b>textClass</b>	<ul style="list-style-type: none"> <li>wissenschaft</li> <li>populaerwissenschaft</li> </ul>	<b>textSigle</b>	ATZ14/FEB/00001	<b>textType</b>	Zeitschrift: Fachzeitschrift	<b>publisher</b>	Springer Fachmedien GmbH	<b>reference</b>	ATZ - Automobiltechnische Zeitschrift, Februar 2014, Nr. 116(2), S. 8-9 - Goppelt, G.: 12. CTI-Getriebesymposium																
<b>textTypeArt</b>	Tagungsbericht	<b>textTypeRef</b>	Fachzeitschrift	<b>title</b>	12. CTI-Getriebesymposium	<b>tokenSource</b>	gingko#morpho	<b>Foundry</b>	Layer	Die	E-Maschine	kann	als	Anfahrelement	dienen	die	Konkurrenz	zwischen	Drehmomentwandler	und	Kupplunge						
corenlp	p	ART	NN	VMFIN	APPR	NN	VVINF	ART	NN	APPR	NN	KON	NN	gingko	l	die	E-Maschine	können	als	Anfahrelement	dienen	die	Konkurrenz	zwischen	Drehmomentwandler	und	Kupplung
gingko	p	ART	NN	VMFIN	KOKOM	NN	VVINF	ART	NN	APPR	NN	KON	NN	marmot	m	case:nom gender:fem number:sg	case:nom gender:fem number:sg	mood:ind number:sg person:3 tense:pres	case:nom gender:neut number:sg	case:nom gender:fem number:sg	case:nom gender:fem number:sg	case:dat gender:masc number:sg	case:nom gender:fem number:pl				
marmot	p	ART	NN	VMFIN	APPR	NN	VVINF	ART	NN	APPR	NN	KON	NN	marmot	p	ART	NN	VMFIN	APPR	NN	VVINF	ART	NN	APPR	NN	KON	NN
opennlp	p	ART	NN	VMFIN	APPR	NN	VVINF	ART	NN	APPR	NN	KON	NN	tt	l	die	E-Maschine	können	als	Anfahrelement	dienen	die	Konkurrenz	zwischen	Drehmomentwandler	und	Kupplung
tt	p	ART	NN	VMFIN	KOKOM	NN	VVFIN	ART	PRELS	APPR	NN	KON	NN														

Abbildung 1: Belegstelle mit Metadaten und Annotationen in KorAP

### 2.1 Texte

Gingko besteht aus  $N=2498$  wissenschaftlichen Artikeln und enthält insgesamt  $N=4.667.656$  Token (Satzzeichen eingeschlossen). Die Artikel stammen aus den beiden auflagenstärksten deutschsprachigen Zeitschriften der Automobil- und Motorentechnik, der Automobiltechnische Zeitschrift (ATZ) und Motortechnische Zeitschrift (MTZ), welche im Springer Professional Verlag mit elf Ausgaben pro Jahr erscheinen. In das Korpus wurden die Jahrgänge 2007 bis 2016 aufgenommen.

Die Zielgruppe der Zeitschriften sind Institutionen aus Forschung und Entwicklung, Unternehmen, Hersteller und Zulieferer, die sich über neue Forschungsergebnisse, Produktions- und Entwicklungsverfahren austauschen. Es handelt sich daher um eine grundsätzlich symmetrische Kommunikation zwischen Expert:innen der Automobiltechnikbranche. Es kann

davon ausgegangen werden, dass die in den Texten verwendete Sprache repräsentativ und normgebend für den Fachbereich ist, da ein großer Teil der Artikel auf Dissertationen und Forschungsprojekten beruht und sie bezogen auf fachliche Inhalte und Sprache begutachtet werden. Gingko enthält eher kürzere Texte mit durchschnittlich 1900 Wörtern mehrerer Autor:innen. Im Gegensatz dazu beinhaltet beispielsweise das wissenschaftliche Subkorpus des DWDS-Kernkorpus 2000-2010 v. a. Monografien. Auch diese Vielfalt steigert die Repräsentativität des Korpus.

Aus Gründen der Datenaufbereitung finden sich in den Jahrgängen 2011 bis 2016 neben der überwiegenden Mehrheit an wissenschaftlichen Artikeln  $N=426$  kürzere Texte mit eher journalistischem Einschlag. Das entspricht insgesamt  $N=249.766$  Token (Satzzeichen eingeschlossen). Es handelt sich zumeist um Meldungen, Interviews, Werbetexte oder andere Texte, die die Zeitschriften gliedern. Diese Texte sind mit der Metadatenvariable `gingkoGenreMain` bzw. `gingkoGenreSub` gekennzeichnet (Kapitel 2.2.2).

In KorAp wird aus Lizenzgründen der Textkontext um den Suchtreffer auf 40 Wörter (Satzzeichen ausgenommen) begrenzt.

## 2.2 Metadaten

In Gingko gibt es Metadaten, die für alle Texte des Korpus übereinstimmen und Metadaten, die für einzelne Texte spezifisch sind. Erstere enthalten z. B. die Information, dass die Sprache des Korpus Deutsch ist; eine textbezogene Information sind beispielsweise die Daten zur Veröffentlichung des Artikels. Die Metadaten sind in KorAP durch Anklicken eines Suchtreffers und das Auswählen der Option „Metadaten“ abrufbar (Kapitel 3.5).

ATZ13/JUL/00006		des Roboterarms nicht <b>vernachlässigbar</b> ist und kompensiert werden muss	
in Summe verkürzt sich die kostenintensive Prüfstandsbelegzeit jedoch erheblich. Zudem werden der Betätigung für zwei Positionen der Wastegateklappe Für bestimmte Positionen der Wastegate Diese gehen mit erhöhten Verschleißgeschwindigkeiten an den turbinenseitigen Dichtringen einhe Rotoraxialbewegungen wurden für das			
<b>author</b>	Sailer, Michael; Hadler, Jens	<b>availability</b>	QAO-NC
<b>corpusSigle</b>	MTZ14	<b>corpusTitle</b>	Gingko - Geschriebenes Ingenieurwissenschaftliches Korpus: MTZ - Motortechnische Zeitschrift, 2014
<b>foundries</b>	<ul style="list-style-type: none"> <li>corenlp</li> <li>corenlp/constituency</li> <li>corenlp/morpho</li> <li>corenlp/sentences</li> </ul>	<b>gingkoArticl...</b>	doi:10.1007/s35146-014-0382-9
<b>gingkoGenr...</b>	wissenschaftlich	<b>gingkoLem...</b>	yes
<b>indexCreati...</b>	2022-03-09	<b>indexLastM...</b>	2022-03-09
<b>pubPlaceKey</b>	DE	<b>publisher</b>	Springer Fachmedien GmbH
<b>textSigle</b>	MTZ14/JUL/00004	<b>textType</b>	Zeitschrift: Fachzeitschrift
<b>tokenSource</b>	gingko#morpho		
<span>Metadaten</span> <span>Token</span> <span>Relationen</span> <b>Werkzeuge für die Komponentenentwicklung von Abgasturbol</b>			

Abbildung 2: Metadaten in Gingko

### 2.2.1 Übersicht über die Metadatenvariablen

In Tabelle 1 werden die wichtigsten Metadatenvariablen erläutert und zum Teil Beispiele für die Werte gegeben.

Variable	Erläuterung
author	Autor:in(nen) des Artikels
availability	Lizenztyp („QAO-NC“ - Query and analysis only-noncommercial: kein Volltextdownload, Nutzende haben das Recht, das Korpus für akademische Zwecke zu durchsuchen und zu analysieren)
corpusEditor	Leiter:innen des Erstellungsprojekts („Prof. Christian Fandrych (Leipzig University); Jun.-Prof. Antje Heine“)
corpusSigle	Kürzel für Name und Publikationsjahr der Zeitschrift (z. B.



Variable	Erläuterung
	„MTZ10“)
corpusExternalLink	Verlinkung zur Website des Gingko-Projekts
corpusTitle	Name des Korpus mit Zeitschrift und Jahr (z. B. „Gingko - Geschriebenes Ingenieurwissenschaftliches Korpus: ATZ - Automobiltechnische Zeitschrift, 2016“)
docSigle	corpusSigle und Monat der Veröffentlichung (z. B. „MTZ10/JUL“)
foundries	Vorhandene Annotationen (Kapitel 2.3)
gingkoArticleDOI	Vollständiger Digital Object Identifier (DOI, z. B. „10.1365/s35148-011-0076-2“)
gingkoCollectionShort	Kürzel des Korpus („Gingko“)
gingkoGenreMain	Textsorte („wissenschaftlich“ oder „journalistisch“)
gingkoGenreSub	Genauere Einteilung der journalistischen Textsorten („wissenschaftlich“ oder alle vorkommenden Werte der Gingko-Variable <b>text-type</b> für journalistische Texte, Kapitel 2.2.2)
gingkoLemmaCorr	„yes“ zeigt an, dass die Lemmaannotation des TreeTaggers ergänzt wurde, wenn der TreeTagger den Wörtern keine Lemmata zuordnen konnte. Dies entspricht den Jahrgängen 2011 bis 2016). Texte mit dem Wert „no“ haben die originale Annotation des TreeTaggers (2007-2010).
gingkoSourceShort	Kürzel des Namens der Zeitschrift („ATZ“ oder „MTZ“)
gingkoTextTokens	Tokenzahl des Textes mit Satzzeichen (z. B. „3334“)
indexCreationDate	Datum der Veröffentlichung von Gingko in DeReKo („2022-03-09“)
indexLastModified	Datum der letzten Modifikation
language	Abkürzung für die Sprache des Artikels („de“)
pubDate	Jahr der Veröffentlichung (z. B. „2014“)
publisher	Herausgeber der Zeitschriften („Springer Fachmedien GmbH“)
pubPlace	Publikationsort der Zeitschriften („Wiesbaden“)
pubPlaceKey	Länderkürzel des Landes der Veröffentlichung („DE“)
reference	Vollständige Literaturangabe des Artikels
srcPages	Seitenzahlen des Artikels (z. B. „516-523“)
textClass	Textklasse, die der DeReKo-Textclassifier ermittelte (z. B.

Variable	Erläuterung
	„technik-industrie“, „kfz“, ...)
textSigle	Identifizier, der im DeReKo über COSMAS II oder KorAP angezeigt wird (z.B. „MTZ11/APR.00001“)
textTypeArt	DeReKo-Texttyp („Fachartikel“)
textTypeRef	DeReKo-Texttyp der Publikation („Fachzeitschrift“)
title	Titel des Artikels
tokenSource	Tokenanzahl des Artikels mit Satzzeichen

*Tabelle 1: Metadatenvariablen und beispielhafte Werte*

## 2.2.2 Informationen zu journalistischen Textsorten

Auch wenn der Verlag dem Projekt vornehmlich wissenschaftliche Texte übergeben hat, wurden während der Projektlaufzeit 426 kürzere Texte mit eher journalistischer Prägung gefunden. Sie machen insgesamt 249.766 Token aus und gehören zu den in den Jahrgängen 2011 bis 2016. Es handelt sich dabei beispielsweise um Meldungen, Interviews und die Zeitschrift gliedernde Texte. Mit der Metadatenvariable `gingkoGenreMain` werden diese 426 Texte als **journalistisch** gekennzeichnet, die restlichen Texte sind als **wissenschaftlich** markiert. Die Metadatenvariable `gingkoGenreSub` ordnet den Texten eine spezifische unterteilte Textsorte zu. Sie wurde händisch, während einer Durchsicht besonders kurzer Artikel, bestimmt. Tabelle 2 erläutert die Textsorten, die als Metadatenwerte gespeichert sind.

Metadatenwert	Erläuterung	Tokenzahl gesamt	Anzahl Texte	Durchschn. Textlänge
Aktuell	Kurzmeldungen zu Entwicklung in der Forschung und Wirtschaft und zu Tagungen	12.006	19	632
Anzeige	Werbung von Springer, Anzeige zu gewonnenem Wettbewerb, Stellenanzeige	1.534	3	511
Bild des Monats	Bild des Monats mit Kurzmeldung zu aktuellen Entwicklungen	2.326	12	194
Bücher	Vorstellung von Buchneuerscheinungen	3.525	7	504
Editorial	Editorial der Chefredaktion zu Inhalten und Entwicklungen der Zeitschrift	3.959	11	360
FVV-Berichte	Meist zwei kurze Berichte der Forschungsvereinigung Verbrennungskraft Maschinen e. V.	4.715	8	589

Metadatenwert	Erläuterung	Tokenzahl gesamt	Anzahl Texte	Durchschn. Textlänge
Gastkommentar	Gastkommentar zu freien Themen am Ende der Zeitschrift	42.276	102	414
Historie	Zusammenfassung der Berichterstattung bezogen auf ein bestimmtes Thema für alle bisher erschienenen Zeitschriften	2.950	5	590
Interview	Interview	20.692	14	1.478
Patente	Vorstellung aktueller Patente	4.171	8	521
Personen + Unternehmen	Kurzmeldungen zu Personen aus der Wirtschaft und Unternehmen	27.638	25	1.106
Porträt	Porträts von Personen aus der Automobilbranche	6.238	12	520
Produkte	Vorstellung neuer Produkte	28.174	26	1.084
Special	Ankündigung eines thematischen Sonderteils mit drei bis sieben zugehörigen Artikeln	400	4	100
Tagungsbericht	Berichte über Tagungen wie die Internationale Automobil-Ausstellung oder die ATZlive-Fachtagung	29.449	28	1.075
Titelthema	Ankündigung des Titelthemas mit ca. drei bis zehn zugehörigen Artikeln	58.170	135	431
Vorschau	Vorstellung des Titelthemas der kommenden Zeitschrift	1.543	7	221

Tabelle 2: Übersicht journalistischer Textsorten

## 2.3 Annotationen

### 2.3.1 Linguistische Annotationen

Für die Texte in Gingko ist eine Vielzahl von Annotationen abrufbar, was die Suche nach bestimmten grammatischen Phänomenen ermöglicht. Sie sind in KorAP als Annotationsschichten und -ebenen gegliedert. Eine Annotationsschicht (*foundry*) benennt die Ressource der Annotationsdaten, z. B. **tt** für die Annotationen, die mit dem *TreeTagger* vorgenommen wurden. Eine Annotationsebene (*layer*) steht für die Art der Annotation: In **tt/p=NN** steht **p** für das Part-of-Speech-Tagging. Bei Suchanfragen ohne Spezifizierung der Ressourcen-Definition (also z. B. nur **p=NN**) wird die als Standard eingestellte Annotationsschicht angesprochen. Bei der Lemma- und POS-Annotation ist die Default-Einstellung die durch das IDS vorgenommene *TreeTagger*-Annotation **tt/1** und **tt/p** (Kupietz et al., 2017, S. 322). Weiterhin unterscheiden sich die Annotationen aus technischer Perspektive dahingehend, welchem Element sie Informationen

zuweisen: Es gibt Annotationen, die sich auf Token (z. B. Lemma oder Part-of-Speech), auf Tokensequenzen (z. B. Konstituenten oder Sätze) und auf Relationen zwischen Token (z. B. Dependenzrelationen) beziehen.

Um die **tokenbezogenen Annotationen** in KorAP anzeigen zu lassen, klickt man in der Trefferansicht auf die Option „Token“ (Kapitel 3.5). Abbildung 3 zeigt die Annotationen des Wortes *Antriebsleistung*. Neben der TreeTragger-Annotation erscheinen Annotationen der Ressourcen *CoreNLP*, *Marmot* und *OpenNLP*. Die Annotationsschicht **gingko** wurde projektintern vor der Übergabe an das IDS erstellt und beinhaltet eine Version der TreeTagger-Annotation<sup>1</sup> (Schmid, 1995; <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>).

Foundry	Layer	Antriebsleistung
corenlp	p	NN
gingko	l	Antriebsleistung
gingko	p	NN
marmot	m	case:nom gender:fem number:sg
marmot	p	NN
opennlp	p	NN
tt	l	Antriebsleistung
tt	p	NN

Metadaten **Token** Relationen Entwicklung Einz...

Abbildung 3: Annotationen eines Tokens

Annotationen, die sich auf **Tokensequenzen** beziehen, sind die Annotationen der Satzspannen **corenlp/s**, **opennlp/s** und **dereko/s**<sup>2</sup> (Abbildung 4) und die der Satzkonstituenten **corenlp/c** (Abbildung 5). In KorAP sind sie gemeinsam mit der **Annotation von Dependenzrelationen malt/d** (Abbildung 6) unter der Option „Relationen“ zusammengefasst. Relationen können nur mithilfe der Suchanfragesprache ANNIS QL abgefragt werden (Kapitel 3.4, Anhang S. 21).

<sup>1</sup> In den Jahrgängen 2011 bis 2016 wurde in der Annotationsebene **gingko/l** ein großer Teil der Lemmata ergänzt, die in der Wortliste des TreeTaggers nicht enthalten waren. Dies geschah, indem die unbekanntenen Wörter automatisch mit einer Wortliste des Dudens abgeglichen und schrittweise verringerter Zuordnungsgenauigkeiten gearbeitet wurden.

<sup>2</sup> **dereko/s** beinhaltet die projektintern erstellte Satzannotation. Satzspannen wurden mittels eines *Python*-Algorithmus aus dem *Natural Language Toolkit* erkannt (Bird et al., 2009; <https://www.nltk.org/>).

Dies bietet Vorteile bezüglich Package und **Antriebsleistung** .

dereko s

Metadaten Token **Relationen** **Geregelte Pendelschieber-Ölpumpe zur bedarfsgerechten Versorgung von Getrieben** von Conrad, Holger; Knauss, Rüdiger (2011) [ATZ11/DEZ/00007]

Abbildung 4: Satzannotation

Druckpumpe zu versorgen. Dies bietet Vorteile bezüglich **Package und Antriebsleistung** .

corenlp c

Metadaten Token Relationen **Geregelte Pendelschieber-Ölpumpe zur bedarfsgerechten Versorgung von Getrieben** von Conrad, Holger; Knauss, Rüdiger (2011) [ATZ11/DEZ/00007]

Abbildung 5: Annotation von Satzkonstituenten

Kühlölkreislauf mit einer bi-dynamischen Druckpumpe zu versorgen. Dies bietet Vorteile bezüglich **Package und Antriebsleistung** .

malt d

Metadaten Token Relationen **Geregelte Pendelschieber-Ölpumpe zur bedarfsgerechten Versorgung von Getrieben** von Conrad, Holger; Knauss, Rüdiger (2011) [ATZ11/DEZ/00007]

Abbildung 6: Annotation von Abhängigkeitsrelationen

Der Vorteil von **alternativen Annotationsschichten** – also mehreren POS-, Lemma- oder Satzannotationen – ist, dass durch Übereinstimmungen oder Abweichungen in den Annotationen

das Trefferergebnis optimiert werden kann (Kupietz et al., 2017, S. 325). Erstens kann man durch die alternative Anfrage zweier Annotationsschichten nicht gefundene Phänomene (falsch Negative) minimieren. Die folgende Suchanfrage sucht *das*, das entweder durch den *TreeTagger* oder durch *CoreNLP* als Relativpronomen getaggt wurde:

```
[orth="das" & (tt/p=PRELS | corenlp/p=PRELS)]
```

Zweitens können in den Treffern aufgeführte Belege minimiert werden, die nicht der Suchintention entsprechen (falsch Positive). Die folgende Suchanfrage sucht *das*, das sowohl durch den *TreeTagger* als auch durch *CoreNLP* als Relativpronomen getaggt wurde:

```
[orth="das" & (tt/p=PRELS & corenlp/p=PRELS)] (Kupietz et al., 2017, S. 325)
```

Eine vollständige Aufschlüsselung der *foundries* bzw. *layers* in KorAP ist in der KorAP-Einführung unter >Korpusdaten>Annotationen auf der KorAP-Plattform abrufbar: <https://korap.ids-mannheim.de/doc/data/annotation>

### 2.3.2 Annotationen zur Dokumentstruktur

Neben den linguistischen Annotationen sind auch Titel, Zwischenüberschriften, Absätze, Bildunterschriften, Tabellen, Formeldarstellungen und Auflistungen in den Artikeln markiert. Die Markierung der Strukturen wurde für die Jahrgänge 2007 bis 2010 händisch vorgenommen, da der Springer-Verlag für diese Jahrgänge nur die PDFs der Artikel bereitstellen konnte. Für die Jahrgänge 2011 bis 2016 wurden XML-Dateien im JATS-Schema (*Journal Publishing Tag Set*) übergeben, sodass die Markierungen der Dokumentstruktur automatisch übernommen werden konnten. Tabelle 3 listet die verfügbaren Annotationen mit ihren Suchausdrücken auf.

Struktur	Suchausdruck in Poliqarp+	Anzahl
Artikeltitel und Zwischenüberschriften	<code>&lt;dereko/s=head&gt;</code>	35.269
Zwischenüberschriften	<code>&lt;dereko/s=head type=sub&gt;</code>	17.766
Absätze	<code>(&lt;dereko/s="p"&gt; &lt;dereko/s=seg type=p&gt;)</code>	66.488
Abbildungsbeschriftungen	<code>&lt;dereko/s=figure&gt;</code>	14.843
Tabellen	<code>&lt;dereko/s=seg type=table-wrap&gt;</code>	814
Formeldarstellungen	<code>&lt;dereko/s=seg type=formula&gt;</code>	1.309
Auflistungen	<code>&lt;dereko/s=list&gt;</code>	1.965

Tabelle 3: Annotationen zur Dokumentstruktur

## 3 Korpusnutzung in KorAP

Gingko wurde in das Deutsche Referenzkorpus (DeReKo, Kupietz et al., 2018) des IDS eingegliedert und kann somit über seine Korpusanalyseplattformen KorAP (Kupietz et al., 2020) und COSMAS II genutzt werden.

Da KorAP COSMAS II in Zukunft ablösen wird, wird im vorliegenden Handbuch nur die Nutzung von Gingko in KorAP beschrieben. Für die Nutzung von COSMAS II bietet die COSMAS-Hilfe Informationen: <https://www2.ids-mannheim.de/cosmas2/projekt/einsteiger/>.

### 3.1 Anmeldung für die IDS-Dienste KorAP und COSMAS II

Bevor KorAP genutzt werden kann (<https://korap.ids-mannheim.de>), muss man sich zunächst für die IDS-Dienste kostenfrei registrieren (<https://perso.ids-mannheim.de/registration/>). Die Anmeldedaten gelten sowohl für KorAP als auch für COSMAS II.

### 3.2 Kennenlernen der Funktionen von KorAP

Auf der KorAP-Startseite befinden sich unter dem Suchfeld und dem Korpusassistenten die Verlinkungen zur Tour und zur Einführung (<https://korap.ids-mannheim.de/doc/>), um die Funktionalitäten von KorAP kennenzulernen (Abbildung 7).

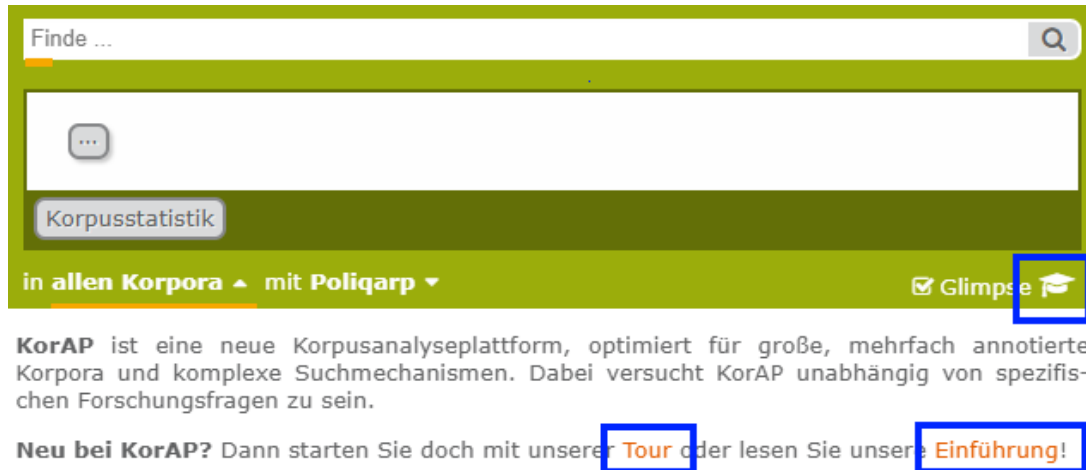


Abbildung 7: KorAP-Tour und Einführung

Zudem kann auf die KorAP-Tutorials zurückgegriffen werden:

[Video 1: Benutzeroberfläche und Korpusauswahl](#)

[Video 2: Suchanfrage, Trefferanzeige sowie die zu den Suchtreffern abrufbaren Informationen](#)

[Video 3: Datenexport und Öffnen in Excel](#)

### 3.3 Korpusassistent: Eingrenzen der Textauswahl, Korpusstatistik, Suchanfragesprache

Um nicht im gesamten DeReKo-Korpus zu suchen, kann zur Definition eines Subkorpus die Textauswahl anhand der Metadaten eingegrenzt werden (Kapitel 2.2). Beispielsweise können alle Artikel eines bestimmten Veröffentlichungsjahres ausgewählt werden, aber auch alle Texte einer Zeitschrift oder die Texte, die einem bestimmten Korpus zugeordnet sind. Im Korpusassistenten kann die Auswahl bestimmter Metadatenvariablen per Klick vorgenommen werden (Abbildung 8).



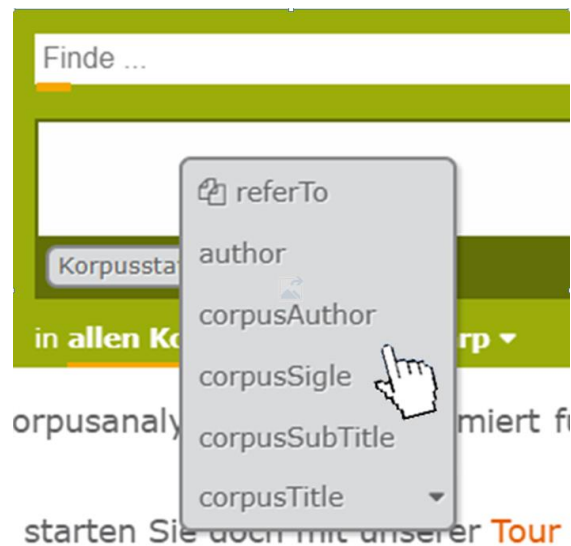


Abbildung 8: Auswahl einer Metadatenvariable

Um die Suche auf die Texte in Gingko einzugrenzen, stehen zwei einfache Wege zur Verfügung:

Als erste Option kann Gingko über die darin enthaltenen Zeitschriften, Automobiltechnischen Zeitschrift (ATZ) und der Motortechnischen Zeitschrift (MTZ), gefunden werden. Dazu muss zunächst aus der Liste der Metadatenvariablen `corpusSigle` ausgewählt werden. Als Werte können `[AM]TZ..`, `[MA]TZ..` oder auch `(ATZ|MTZ)..` eingegeben werden und das Kästchen mit „RegEx“ muss aktiviert werden (Abbildung 9).



Abbildung 9: Auswahl von Gingko über „corpusSigle“

Als zweite Option kann `gingkoCollectionShort` direkt eingegeben und durch die Enter-Taste bestätigt werden, statt eine Metadatenvariable aus der gegebenen Liste auszuwählen. Dann klickt man die drei Punkte rechts neben dem erscheinenden Kästchen „eq“ an und gibt den Wert `Gingko` ein. Hierbei ist es wichtig, die Groß- und Kleinschreibung zu beachten, da Gingko sonst nicht gefunden wird (Abbildung 10).

`gingkoCollectionShort eq Gingko`

Abbildung 10: Auswahl von Gingko über „gingkoCollectionShort“

Über den Button „**Korpusstatistik**“ wird die Anzahl der Dokumente, Paragraphen, Sätze und Token in einem Korpus angezeigt. Die Tokenanzahl schließt dabei die Satzzeichen aus. Nach der vorgenommenen Korpuseingrenzung sieht man die Zahlen für Gingko (Abbildung 11).

corpusSigle eq /[AM]TZ../

documents:	2.498	paragraphs:	51.919
sentences:	225.029	tokens:	4.123.158
Korpusstatistik			

Abbildung 11: Korpusstatistik

Weiterhin wird über den Korpusassistenten auch die **Suchanfragesprache** ausgewählt (Abbildung 12). Die meisten Beispielsuchanfragen im Anhang (S. 21) sind mit Poliqarp+ formuliert, da diese Sprache die meisten Funktionalitäten bietet.

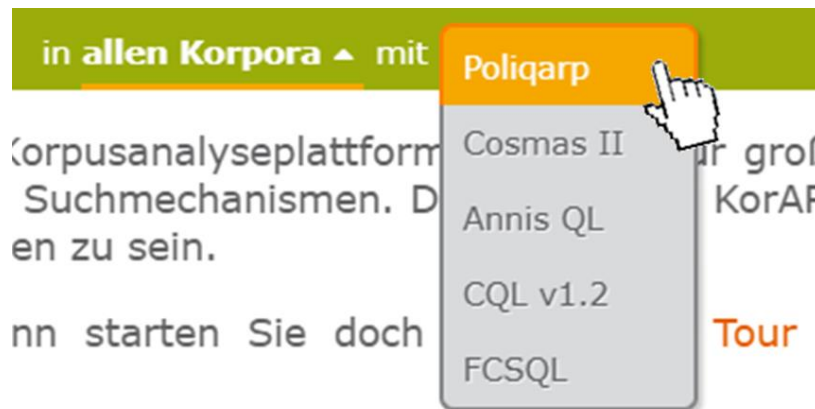


Abbildung 12: Auswahl der Suchanfragesprache

### 3.4 Stellen einer Suchanfrage

Es ist möglich, ein beliebiges Wort oder eine beliebige Wortfolge in das Suchfeld einzugeben, ohne eine Suchanfragesprache explizit zu verwenden. Außerdem besteht die Möglichkeit, eine Beispielabfrage aus der Tabelle im Anhang (S. 21) anzuklicken oder durch Kopieren in das Suchfeld in KorAP einzufügen. Zudem können die in den grauen Kästen hinterlegten Beispielsuchanfragen in der KorAP-Einführung (<https://korap.ids-mannheim.de/doc/>) durch Anklicken auch direkt in das Suchfeld übernommen werden (Abbildung 13). Eine Suche wird stets durch das Anklicken der Lupe oder das Drücken der Enter-Taste ausgeführt.

Anfragesprachen
Cosmas II
Poliqarp+
Einfache Segmente
Komplexe Segmente
Span-Segmente
Paradigmatische Operatoren
Syntagmatische Operatoren
Klassen-Operatoren
Annis QL
CQL

[base=Baum]

Der dritte Sonderschlüssel ist `pos` und bezieht sich auf die

[pos=ADJA]

Komplexe Segmente, die weitere Token-Annotationen erlauben. Das Vorkommen von Plural-Wörtern in einer dieser Informationen

[mate/m=number:pl]

Falls eine Annotation Sonderzeichen enthält, sorgen umschlüsselnde

[orth='http://www.ids-mannheim.de/cosmas2/projekt/']

Abbildung 13: Übernehmen eines Suchausdrucks aus der KorAP-Einführung

### 3.5 Trefferansicht

Nachdem eine Suche ausgeführt wurde, werden alle Ergebnisse als KWIC dargestellt, sodass die Treffer zentriert untereinander erscheinen. Durch das Anklicken eines Belegs wird dieser ausgewählt, erscheint orange und weiterer Kontext wird angezeigt (Abbildung 14). Zudem wird rechts oben die Trefferzahl ausgegeben und rechts unten kann zu weiteren Seiten der Suchtreffer navigiert werden. Die Anzahl der Treffer wird wie in Abbildung 14 dann angegeben, wenn die Glimpse-Funktion deaktiviert ist. Ist diese Funktion aktiviert, läuft die Suche schneller ab, die Ergebnisse werden jedoch auf die erste Seite reduziert und die Trefferzahl wird nicht angezeigt.

Wenn ein Treffer durch Anklicken ausgewählt wurde, werden links unten die grau hinterlegten Buttons „Metadaten“, „Token“ und „Relationen“ sichtbar, über die man zu den Metadaten und Annotationen der Belegstelle gelangt (Kapitel 2.2 und 2.3, Abbildung 15).

The screenshot shows the KorAP search interface. The search query is 'vernachlässigbar'. The results are displayed in a KWIC format. The top result is highlighted in orange. On the right side, there is a 'Glimpse' button and a '65 Treffer' indicator. At the bottom right, there are navigation buttons for page 1, 2, 3, and a search icon.

Abbildung 14: Trefferanzeige ohne Glimpse-Funktion

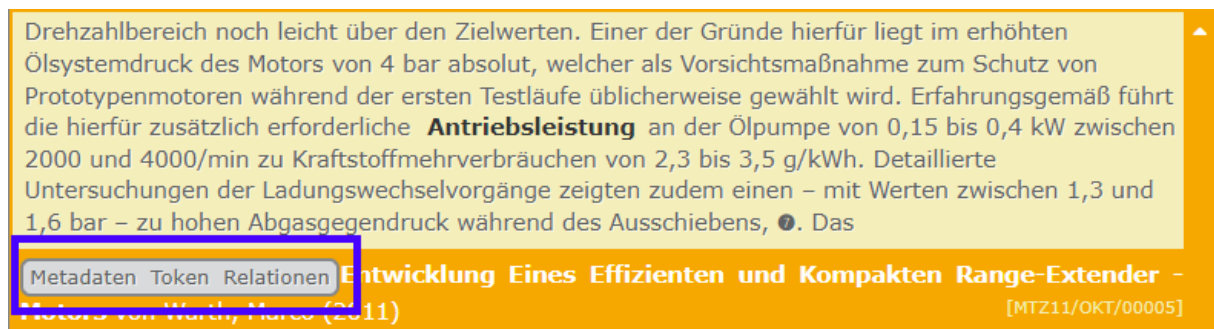


Abbildung 15: Metadaten und Annotationen eines Suchtreffers

### 3.6 Abfragetricks – Zusammenstellen einer Suchanfrage über Trefferansicht

Die in der Trefferansicht gezeigten Metadaten und tokenbezogenen Annotationen eines Suchtreffers können unmittelbar zur Formulierung einer neuen Suchanfrage verwendet werden. Um die Metadaten zur Eingrenzung einer neuen Textauswahl bzw. Annotationen für eine neue Suchanfrage auszuwählen, klickt man die gewünschten Werte an. Dann erscheinen diese grün hinterlegt und werden als neue Bedingung bzw. neue Anfrage unterhalb der Metadaten- bzw. Annotationsanzeige innerhalb eines grauen Fensters angezeigt: Siehe Abbildung 16 für das Spezifizieren der Metadaten und Abbildung 17 für das Festlegen der Annotationen. Durch ein weiteres Anklicken des grauen Fensters werden die ausgewählten Metadaten bzw. Annotationen oben in den Korpusassistenten bzw. in das Suchfeld übernommen. Eine neue Abfrage mit diesen Parametern kann über die Lupe oder die Enter-Taste gestartet werden. Es können auch gleich mehrere Metadatenwerte oder mehrere Annotationen angewählt und kombiniert werden. Dependenzrelationen (Kapitel 2.3.1) können nicht per Klick, sondern nur über die manuelle Eingabe ins Suchfeld mithilfe der Anfragesprache ANNIS QL abgefragt werden.

gingkoColle...	Gingko - Geschriebenes Ingenieurwissenschaftliches Korpus	gingkoColle...	Gingko
gingkoGenr...	wissenschaftlich	gingkoGenr...	wissenschaftlich
gingkoLem...	yes	gingkoSource	MTZ - Motortechnische Zeitschrift
gingkoSour...	MTZ	gingkoTextT...	2257
indexCreati...	2022-03-09	indexLastM...	2022-03-09
language	de	pubDate	2011
pubPlace	Wiesbaden	pubPlaceKey	DE
publisher	Springer Fachmedien GmbH	reference	MTZ - Motortechnische Zeitschrift, Oktober 2011, Nr. 72(10), S. 738-747 - Warth, M.: Entwicklung Eines Effizienten und Kompakten
srcPages	738-747	textClass	<ul style="list-style-type: none"> <li>wissenschaft</li> <li>populaerwissenschaft</li> </ul>
textSigle	MTZ11/OKT/00005	textType	Zeitschrift: Fachzeitschrift
textTypeArt	Fachartikel	textTypeRef	Fachzeitschrift
title	Entwicklung Eines Effizienten und Kompakten Range-Extender -Moto...	tokenSource	gingko#morpho

Neue Bedingung:  
gingkoGenreMain eq wissenschaftlich

Metadaten Token Relationen Entwicklung Eines Effizienten und Kompakten Range-Extender  
Motors von Warth, Marco (2011) [MTZ11/OKT/00005]

Abbildung 16: Neue Suchbedingung über Auswahl der Metadaten

Foundry	Layer	Antriebsleistung	an	der	Ölpumpe	von	0,15	bis	x
corenlp	p	NN	APPR	ART	NN	APPR	CARD	APPR	
gingko	l	Antriebsleistung	an	die		von	@card@	bis	
gingko	p	NN	APPR	ART	NN	APPR	CARD	KON	
marmot	m	case:nom gender:fem number:sg		case:dat gender:fem number:sg	case:dat gender:fem number:sg				
marmot	p	NN	APPR	ART	NN	APPR	CARD	APPR	
opennlp	p	NN	APPR	ART	NN	APPR	CARD	APPR	
tt	l	Antriebsleistung	an	die	Ölpumpe	von	@card@	bis	
tt	p	NN	APPR	ART	NN	APPR	CARD	APPR	KON

Neue Anfrage: [tt/p=APPR][tt/p=ART][marmot/m=case:dat  
tt/p=NN]

Metadaten Token Relationen Entwicklung Eines Effizienten und Kompakten Range-Extender  
Motors von Warth, Marco (2011) [MTZ11/OKT/00005]

Abbildung 17: Neue Suchanfrage über Auswahl der Tokenannotationen



### 3.7 Datenexport

Die Ergebnisse einer Suchanfrage können als KWIC aus KorAP exportiert werden. Hierzu muss der Export-Button mit dem Download-Pfeil rechts neben der Trefferanzahl angeklickt werden (Abbildung 18).



Abbildung 18: Ergebnisexport

## 4 Zitieren des Korpus

Soll das Korpus als Ganzes zitiert oder auf dieses verwiesen werden, dient der folgende Artikel als Quelle:

Schirrmeister, Lars; Rummel, Marlene; Heine, Antje; Suppus, Nina & Mendoza Sánchez, Bárbara (2021). Gingko – ein Korpus der ingenieurwissenschaftlichen Sprache. In: *Deutsch als Fremdsprache*, 58/4, 214–224. <https://doi.org/10.37307/j.2198-2430.2021.04.04>

Belege aus Gingko sollten mit der Originalquelle, also der Zeitschrift, in der sie erschienen sind, zitiert werden, z. B.:

MTZ10/OKT/00005, Motortechnische Zeitschrift, Oktober 2010, Nr. 71(10), S. 694–698, Koch, A.; Claus, H.; Frankenstein, D.; Herfurth, R.: Turboladerdesign für elektrische Waste-Gate-Betätigung minimiert Leckage, DOI:10.1007/BF03225612

Diese Information kann aus KorAP unter der Metadatenvariable **reference** entnommen werden (Kapitel 2.2 und 3.5).

## Literaturverzeichnis

- Berlin-Brandenburgische Akademie der Wissenschaften. (o. J.). *DWDS – Digitales Wörterbuch der deutschen Sprache: Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart*.  
<https://www.dwds.de>
- Heine, A. (2019). Wozu (ge-)brauchen Ingenieure figurative Verben? Eine Analyse auf Basis wissenschaftlicher Artikel. *Informationen Deutsch als Fremdsprache*, 47(6), 612–629.  
<https://doi.org/10.1515/infodaf-2019-0057>
- Heine, A. & Rummel, M. (2019). Funktion statt Transformation: Kausale Relationen in wissenschaftlichen Texten. In I.-A. Busch-Lauer (Hrsg.), *DaFF-Impulse – Deutsch als Fremd- und Fachsprache an Hochschulen* (S. 11–32). Frank & Timme.
- Heine, A., Schirrmeyer, L. & Anderson, M. O. (2019). The Better the Description, the Better the Instruction: Passives and Impersonals in the German Language of (Automotive) Engineering. *German as a foreign language*(2). <http://www.gfl-journal.de/2-2019/heine.pdf>
- Hirschmann, H. (2019). *Korpuslinguistik: Eine Einführung*. J.B. Metzler'sche Verlagsbuchhandlung & Carl Ernst Poeschel GmbH.
- Kupietz, M., Diewald, N., Hanl, M. & Margaretha, E. (2017). Möglichkeiten der Erforschung grammatischer Variation mithilfe von KorAP. In M. Konopka & A. Wöllstein (Hrsg.), *Grammatische Variation: Empirische Zugänge und theoretische Modellierung* (S. 319–330). De Gruyter. <https://doi.org/10.1515/9783110518214-019>
- Kupietz, M., Diewald, N., Margaretha, E., Bodmer, F., Stallkamp, H. & Harders, P. (2020). Recherche in Social-Media-Korpora mit KorAP. In K. Marx, H. Lobin & A. Schmidt (Hrsg.), *Deutsch in Sozialen Medien* (S. 373–378). De Gruyter.  
<https://doi.org/10.1515/9783110679885-024>
- Kupietz, M., Lungen, H., Kamocki, P. & Witt, A. (2018). The German Reference Corpus DeReKo: New Developments – New Opportunities. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis & T. Tokunaga (Hrsg.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (S. 4353-4360). European Language Resources Association (ELRA).
- Meißner, C. (2022). Attributrealisierung und Nominalstil in der Sprache der Ingenieurwissenschaften: Kollexemanalysen zum Gebrauch von Partizipialattributen und Relativsätzen. *Deutsch als Fremdsprache*, 59(2), 88–104.
- Rummel, M. & Heine, A. (2019). Von lexikalischen Markern zu komplexen Mustern – zum semantischen Profil automobiltechnischer Fachtexte. *Fachsprache*, 41(3-4), 123–141.  
<https://doi.org/10.24989/fs.v41i3-4.1726>
- Schmid, H. (1995). Improvements in Part-of-Speech Tagging with an Application to German. In *Proceedings of the ACL SIGDAT-Workshop*.

## Anhang: Beispielsuchanfragen für Gingko in KorAP

Phänomen	Suchausdruck in Poliqarp+	Trefferzahl	Beispiele	Fehltreffer
beliebiges Wort wie <i>luftvergütet</i>	<a href="#">luftvergütet</a> die Form <i>luftvergütet</i>	5	<ul style="list-style-type: none"> <li>• Bild 4: Gefügeausbildung des Stahls BAS100 für verschiedene Anlasszu-stände – weichgeglüht (oben) und <b>luftvergütet</b> (unten).</li> </ul>	

### Substantiv

Phänomen	Suchausdruck in Poliqarp+	Trefferzahl	Beispiele	Fehltreffer
substantivische Derivate	<a href="#">[l=".*heit"]</a> ein Lemma, das auf <i>-heit</i> endet	5.773	<ul style="list-style-type: none"> <li>• Mehrheit</li> <li>• Treffsicherheit</li> <li>• Radantriebseinheit</li> </ul>	
	alternativ in COSMAS II: <a href="#">*heit</a>	4.987		
Konversion zu Substantiven	<a href="#">[p=ART &amp; orth="d(as em es)"]</a> <a href="#">[p=NN &amp; orth="Um.*e(n ns)"]</a>	52	<ul style="list-style-type: none"> <li>• Nach <b>dem Umströmen</b> der KAS-Innenseite gelangt die Luft [...]</li> <li>• [...] während <b>des Umschaltens</b> vom Bypass- auf Kühlerbetrieb.</li> </ul>	Andere Wörter, die auf <i>Um-</i> beginnen und <i>-en(s)</i> enden und keine Konversionen sind, z. B. ... <b>des</b>



	ein definiter Artikel mit den Endungen für Neutrum Singular (-as/-em/-es) gefolgt von einem Substantiv, das mit <i>Um-</i> beginnt und auf <i>-en(s)</i> endet			<i>Umweltgedankens ... , ... das Umsatzverhalten ...</i>
Komposita – Trunkierung	<a href="#">[p=TRUNC]{1,}</a> <a href="#">[p=KON]</a> <a href="#">[p=NN]</a>  eine oder mehrere aufeinanderfolgende Trunkierungen gefolgt von einer Konjunktion und einem Substantiv; nur Treffer für Komposita mit geteiltem Determinatum, nach Trunkierungen mit geteiltem Determinans wird nicht gesucht	14.792	<ul style="list-style-type: none"> <li>[...] die <b>Werkstoff-, Gestalt-, Konzept- und Fertigungsleichtbau</b> als Voraussetzung für zielführenden Bedingungsleichtbau vereinen.</li> <li>Je größer der <b>Substitutions- beziehungsweise Integrationsumfang</b> bei Verbundgusstrukturen ist, [...]</li> </ul>	

## Verb

Phänomen	Suchausdruck in Poliqarp+	Trefferzahl	Beispiele	Fehltreffer
Präteritum	<a href="#">[p=VVFIN &amp; marmot/m=tense:past &amp; marmot/m=mood:ind]</a>	11.977	<ul style="list-style-type: none"> <li>Hierbei <b>hielten</b> sich die Projektmitarbeiter streng an das objektivierte Szenario.</li> <li>Um das CO2-Ziel der Europäischen Automobilindustrie zu erreichen, <b>trug</b> der Dieselmotor mit Common-Rail-Direkteinspritzung und Turboaufladung wesentlich bei.</li> </ul>	als finites Verb getaggte Token im Konjunktiv, z. B. <i>Nach wie vor <b>böte</b> der Verbrennungsmotor das größte Potenzial ...</i> und als finites Verb getaggte Token, die Adjektive sind,

	eine finite Verbform, die als Vergangenheitsform im Indikativ getaggt ist			z. B. ... <i>die aus Endkundensicht <b>definierten</b> Entwicklungsziele ...</i>
Präteritum von <i>bieten</i>	<a href="#">[l=bieten &amp; marmot/m=tense:past &amp; marmot/m=mood:ind]</a>  das Lemma <i>bieten</i> , das als Vergangenheitsform im Indikativ getaggt ist	37	<ul style="list-style-type: none"> <li>• Dieses Verfahren <b>bot</b> schnelle Rückmeldung an die Konstrukteure [...]</li> <li>• Hierfür <b>boten</b> sich systematische Untersuchungen von Ventilsteuerzeiten [...]</li> </ul>	als <i>bieten</i> lemmatisierte Token im Konjunktiv, z. B. <i>Nach wie vor <b>böte</b> der Verbrennungsmotor das größte Potenzial ...</i>
Konjunktiv II	<a href="#">[p=VVFIN &amp; marmot/m=tense:past &amp; marmot/m=mood:subj]</a>  eine finite Verbform, die als Vergangenheitsform im Konjunktiv getaggt ist	314	<ul style="list-style-type: none"> <li>• Ein Messgitter mit 20 Mikrofonen pro Richtung <b>bestünde</b> bereits aus 400 Mikrofonen [...]</li> <li>• Alternativ <b>stünden</b> auch andere Werkstoffe zur Verfügung, wenn es der Kunde wünscht.</li> </ul>	als Konjunktiv getaggte Token im Indikativ Präteritum, z. B. ... <i><b>führen</b> zu höheren mechanischen und insbesondere thermischen Belastungen ...</i>
Verb mit Präpositionalobjekt	<a href="#">[p=VVFIN][l=über][marmot/m=case:acc]{1,3}[p=NN]</a>  ein finites Verb gefolgt vom Lemma <i>über</i> , einer bis drei Akkusativ-Ergänzungen und einem Substantiv	806	<ul style="list-style-type: none"> <li>• [...] <b>verfügt über</b> ein nachgeschaltetes Volumenstromregelventil.</li> <li>• [...] <b>erfolgt über</b> eine Axialverschiebung.</li> <li>• Dies <b>geschieht über</b> ein Mobilfunknetz.</li> </ul>	Abfrage geht über die Satzgrenze hinaus, z. B. ... <i>oben gelegene Reservoir <b>fördert. Über eine Leitung</b> gelangt ...</i>
	alternativ in <b>Annis QL</b> : <a href="#">[l=über/-&gt;maIt/d[func="OBJP"] p=VVFIN/</a>	853	<ul style="list-style-type: none"> <li>• Er <b>verfügt über</b> einen Halter zur Befestigung am Systemträger.</li> <li>• Konventionelle hydraulische Bremsen <b>verfügen über</b> eine mechanisch-hydraulische Rückfallebene.</li> </ul>	als <i>über</i> lemmatisierte Token, die keine Komplemente des Verbs einleiten, sondern beispielsweise Attribute, z. B. <i>Ein unterhalb des</i>

das Lemma <i>über</i> ist als Präpositionalobjekt mit einer finiten Verbform verknüpft, wobei die Reihenfolge der Token beliebig ist			<i>Zylinderkopfs applizierbares Drall und Tumble-Messgerät <b>liefert</b> Messwerte <b>über</b> die Verwirbelung ...</i>
--	--	--	--

## Adjektiv

Phänomen	Suchausdruck in Poliqarp+	Trefferzahl	Beispiele	Fehltreffer
Partizip I attributiv	<a href="#">[l=".*nd" &amp; p=ADJA]</a>  ein Lemma, das auf <i>-nd</i> endet und zugleich als Adjektiv getaggt ist	25.401	<ul style="list-style-type: none"> <li>Die Summe der <b>beanspruchenden</b> Lastspiele steigt [...]</li> <li>Darüber hinaus ist ersichtlich, dass die Eindringtiefen auch bei <b>variierendem</b> Einspritzdruck [...]</li> </ul>	andere Token, die auf <i>-nd</i> enden und keine Partizipialattribute sind, z. B. <i>Das Diagnosesystem lernt deshalb zum Versuchsbeginn das typische, „gesunde“ Verhalten des Prüflings ...</i>
Partizip II attributiv	<a href="#">[l=".*ge.*(t en)" &amp; p=ADJA]</a>  ein Lemma, das <i>-ge-</i> enthält und auf <i>-t</i> bzw. <i>-en</i> endet und zugleich als Adjektiv getaggt ist; nach Präfixverben und Verben auf <i>-ieren</i> wird so nicht gesucht	34.741	<ul style="list-style-type: none"> <li>[...] Oktanbedarf bei modernen <b>aufgeladenen</b> Ottomotoren [...]</li> <li>Für eine aussagekräftige Lebensdauersimulation sind <b>abgesicherte</b> Spannungsberechnungen [...]</li> </ul>	andere Token, die auf <i>-t</i> oder <i>-en</i> enden und keine Partizipialattribute sind, z. B. <b>gesamt, bekannt, verschiedene</b>
Gerundiv (zu + Partizip I)	<a href="#">[orth="zu" &amp; p=PTKZU][l=".*nd" &amp; p=ADJA]</a>	979	<ul style="list-style-type: none"> <li>[...] im Hinblick auf die <b>zu berücksichtigende</b> Kraftstoffqualität [...]</li> <li>Ein nicht <b>zu unterschätzendes</b> Potenzial [...]</li> </ul>	andere Token nach <i>zu</i> , die keine Partizipialattribute sind, z. B. <i>Das Auswahlverfahren bei Delphi führte</i>

	zu gefolgt von einem Lemma, das auf <i>-nd</i> endet und zugleich als Adjektiv getaggt ist			<i>letztlich zu folgender Systemkombination...</i>
--	--	--	--	--

## Präposition

Phänomen	Suchausdruck in Poliqarp+	Trefferzahl	Beispiele	Fehlreffer
Wechselpräposition	<a href="#">[l=über] [marmot/m=case:a cc]</a> das Lemma <i>über</i> gefolgt von einer Akkusativ-Ergänzung	9.939	<ul style="list-style-type: none"> <li>[...] Einzelteilerfertigung des Hybriden <b>über die</b> Fügefolgen im Rohbau [...]</li> <li>Ein reales Testfahrzeug detektiert <b>über virtuelle</b> Sensoren simulierte parkende Fahrzeuge.</li> </ul>	
	<a href="#">[l=über] [marmot/m=case:d at]</a> das Lemma <i>über</i> gefolgt von einer Dativ-Ergänzung	1.090	<ul style="list-style-type: none"> <li><b>Über der</b> Vorderachse sind E-Motor und Getriebe montiert.</li> <li>[...] mit ihrer <b>über 130-jährigen</b> Erfahrung im Automobilbau [...]</li> </ul>	als Dativ getaggte Token im Akkusativ, z. B. <i>Im Motorbetrieb wandelt die Leistungselektronik den Gleichstrom <b>über Hochleistungstransistoren</b> ...</i>

## Konjunktion

Phänomen	Suchausdruck in Poliqarp+	Trefferzahl	Beispiele	Fehltreffer
doppelte Konjunktionen	<a href="#">[orth="sowohl"][]f0,}[orth="als"]</a> <a href="#">[orth="auch"]</a>  <i>sowohl</i> gefolgt von keinem oder beliebig vielen Token und <i>als auch</i>	2.547	<ul style="list-style-type: none"> <li>Der Motor ist <b>sowohl</b> im Brennraum <b>als auch</b> ein- und auslassseitig [...]</li> <li>Hiermit lassen sich <b>sowohl</b> die Potenziale bezüglich Verbrauch und Emissionen eines solchen Downsizingkonzepts ermitteln, <b>als auch</b> dessen Grenzen aufzeigen.</li> </ul>	
	<a href="#">[orth="weder"][]f0,}[orth="noch"]</a>  <i>weder</i> gefolgt von keinem oder beliebig vielen Token und <i>noch</i>	182	<ul style="list-style-type: none"> <li>[...] Schaltvorgänge <b>weder</b> den subjektiven Fahrindruck <b>noch</b> den Fahrkomfort des Fahrzeugs beeinträchtigen.</li> <li>Der Kühlkreislauf ist <b>weder</b> thermostat- <b>noch</b> durchflussgeregelt.</li> </ul>	
	<a href="#">[orth="je"][]f1,}[orth="desto"]</a>  <i>je</i> gefolgt von einem oder mehr beliebigen Token und <i>desto</i>	41	<ul style="list-style-type: none"> <li>Je länger der Zyklus andauert und <b>je</b> früher der Motorwarmlauf beendet ist, <b>desto</b> höher fallen [...].</li> <li>[...] <b>je</b> schneller die Handbewegungen, <b>desto</b> höher ist die Scrollgeschwindigkeit.</li> </ul>	Abfrage geht über die Satzgrenze hinaus, z. B. ... <i>zwar je nach Art des Atomkerns (chemisches Element) und der Energie des Neutrons, ganz unterschiedlich. Für Neutronen gilt nicht wie für Röntgen, je schwerer ein Atom, desto stärker ...</i>

## Relativsatz

Phänomen	Suchausdruck in Poliqarp+	Trefferzahl	Beispiele	Fehltreffer
Relativsatz	<a href="#">[p=PRELS]</a>  ein Relativpronomen	32.330	<ul style="list-style-type: none"> <li>Zulieferer gelten als Partner, <b>die</b> ein automobiles Subsystem ganzheitlich überblicken.</li> <li>Die Querbeschleunigungen, <b>welche</b> zum Beispiel durch Kurven hervorgerufen werden, [...]</li> </ul>	andere Token, die als Relativpronomen getaggt sind, z. B. ... <i>Reduzierung der Betriebskosten, <b>das</b> heißt, für den Transport der gleichen Gütermenge wird weniger Kraftstoff verbraucht.</i>

## Passivsatz

Phänomen	Suchausdruck in Poliqarp+	Trefferzahl	Beispiele	Fehltreffer
<i>sich lassen</i> + Infinitiv	<a href="#">matches(contains(&lt;base/s=s&gt;,[l=lassen]),contains(&lt;base/s=s&gt;,[p=PRF]))</a>  das Lemma <i>lassen</i> und ein reflexives Personalpronomen innerhalb eines Satzes, wobei die Reihenfolge der Token beliebig ist	85	<ul style="list-style-type: none"> <li>[...] mit denen <b>sich</b> Ethanolkraftstoff wirtschaftlich aus Agrar-Reststoffen <b>herstellen lässt</b>.</li> <li>Ferner <b>lässt sich</b> der Gesamtwirkungsgrad des elektrischen Verdichters in den Positionen 3 und 4 durch die Möglichkeit, die Verdichterkennfelder schmaler auszuführen, positiv <b>beeinflussen</b>.</li> </ul>	das Pronomen <i>sich</i> gehört nicht zur Passivparaphrase, z. B. <i>Statt die störenden Signale <b>sich</b> mit hohen Spannungsanstiegsgeschwindigkeiten über die Leitungen ausbreiten zu <b>lassen</b>, ...</i>

	<p>alternativ in <b>Annis-QL</b>:  <a href="#">corenlp/s="s" &amp; tt/l="lassen" &amp; l="sich" &amp; #1_i_#2 &amp; #1_i_#3</a></p> <p>das Lemma <i>lassen</i> und das Lemma <i>sich</i> innerhalb eines Satzes, wobei die Reihenfolge der Token beliebig ist</p>	4.539	<ul style="list-style-type: none"> <li>• Kamerasysteme <b>lassen sich</b> aber auch für Mauterfassung <b>auslegen</b>.</li> <li>• [...] wodurch <b>sich</b> hohe Leistungsdichten, geringes Einbauvolumen und -gewicht sowie Kostenersparnisse <b>erzielen lassen</b>.</li> </ul>	
<p><i>sein</i> + <i>zu</i> + Infinitiv</p>	<p><a href="#">[l=sein][l{,3}[p=PTKZU][p=VVINF]</a></p> <p>das Lemma <i>sein</i> gefolgt von bis zu drei beliebigen Token, dem Partikel <i>zu</i> und einem Infinitiv</p>	1.435	<ul style="list-style-type: none"> <li>• Die HCCI-Selbstzündung <b>ist</b> bedeutend schwieriger <b>zu kontrollieren</b>.</li> <li>• Deutlich <b>ist</b> den Diagrammen <b>zu entnehmen</b>, dass selbst Verbrennungsmotoren [...]</li> <li>• Solche Betriebszustände <b>sind</b> daher <b>zu vermeiden</b>.</li> </ul>	<p>das Token <i>zu</i> gefolgt von Infinitiv als Teil eines Nebensatzes mit Infinitivgruppe, z. B. ... <i>sehr kompakt und effizient sein, um Fahrspaß zu erzeugen</i>; und Abfrage geht über die Satzgrenze hinaus, z. B. ... <i>eines Türkonzepts sein. Wie zu erwarten wurde ...</i></p>
<p><i>sein</i> + Adjektiv auf <i>-bar/-lich/-abel</i></p>	<p><a href="#">[l=sein][l{,3}[p=ADJD &amp;(orth=".*bar" orth=".*lich" orth=".*abel")]</a></p> <p>das Lemma <i>sein</i> gefolgt von bis zu drei beliebigen Token und einem Adjektiv, das auf <i>-bar</i> oder <i>-lich</i> oder <i>-abel</i> endet</p>	3.772	<ul style="list-style-type: none"> <li>• Der bainitische Stahl <b>ist</b> schwerer <b>bearbeitbar</b>.</li> <li>• Hier <b>sind</b> deutlich differenziertere Konzepte <b>erforderlich</b>.</li> <li>• Allein dadurch <b>ist</b> es <b>möglich</b>, in bestimmten Situationen [...]</li> <li>• Es <b>ist</b> nicht immer <b>praktikabel</b>, einen minimalen Ausgleichsbehälterdruck zu gewährleisten.</li> </ul>	<p>das Token <i>sein</i> gefolgt von als Adjektiv getaggten Token, die auf <i>-bar/-lich/-abel</i> enden und keine Passivparaphrase bilden, z. B. <i>Die berechneten Konzentrationen [...] sind deutlich höher als die gemessenen</i>; und Abfrage geht über die Satzgrenze hinaus, z. B. ... <i>individualisierbar sein. Ferner müssen Leuchten dimmbar sein ...</i></p>

<p><i>müssen/ können/ sollen + werden +</i> Partizip II</p>	<p><a href="#">[l="müssen" l="können" l="sollen"]{0,5}[p=VVPP]</a> <a href="#">][l="werden"]</a></p> <p>das Lemma <i>müssen, können</i> oder <i>sollen</i> gefolgt von keinem oder bis zu fünf Token, einem Partizip Perfekt und dem Lemma <i>werden</i></p>	<p>6.615</p>	<ul style="list-style-type: none"> <li>• Daher <b>sollen</b> adaptive Regler <b>getestet werden</b>, die [...]</li> <li>• Kostspielige Iterationsschleifen im Freigabeprozess <b>können</b> damit <b>vermieden werden</b>.</li> </ul>	<p>Abfrage geht über die Satzgrenze hinaus, z. B. ... <i>damit sich die Schwingungsformen ausbilden können</i>. <b>Identifiziert wurde</b> die erste Torsionseigenfrequenz ...</p>
---	--	--------------	---	--