

GINGKO

Geschriebenes ingenieurwissenschaftliches Korpus

Annette Portmann
Universität Leipzig

Abstract

Das Geschriebene ingenieurwissenschaftliche Korpus *Gingko* ist im *Deutschen Referenzkorpus* (DeReKo; vgl. Kupietz et al. 2018) am Leibniz-Institut für Deutsche Sprache (IDS) veröffentlicht und über die IDS-Korpusplattformen frei zugänglich. Es ermöglicht Korpusanalysen, die Empfehlungen für relevante Unterrichtsinhalte in universitären Sprachkursen spezifisch für technische Studiengänge empirisch fundieren können. Für Studierende und für Lehrende bietet das Korpus eine Ressource für Beispiele von Wörtern und Strukturen im Kontext von Texten, die als angestrebte Norm betrachtet werden können.

Keywords: Korpus; ingenieurwissenschaftliche Sprache; Deutsch als fremde Wissenschaftssprache; Fachsprache

Abstract

The Written Corpus for the Academic Language of Engineering *Gingko* is published in the *German Reference Corpus* (DeReKo; cfr. Kupietz et al. 2018) at the Leibniz-Institut für Deutsche Sprache (IDS) and can be accessed online via the corpus research platforms of the IDS. It allows corpus analyses which can substantiate recommendations for contents in specialized academic language courses. Students and language teachers in technical study programmes are thus given a resource that enables them to search for typical usages of selected lexical items and structures which are part of texts that can be considered a target standard.

Keywords: corpus; language of engineering; German as a foreign academic language; specialized language

Gingko ist ein Korpus der geschriebenen ingenieurwissenschaftlichen Fachsprache (vgl. Schirmmeister et al. 2021). Es ist im *Deutschen Referenzkorpus* (DeReKo; vgl. Kupietz et al. 2018) am Leibniz-Institut für Deutsche Sprache (IDS) veröffentlicht und über die Korpusplattformen *KorAP* (vgl. Kupietz et al. 2020) und *COSMAS II* abrufbar. Im Folgenden wird ein Überblick über die Relevanz der Daten und das Korpusdesign gegeben. Anschließend soll eine kurze korpusvergleichende Analyse zu figurativen Lesarten des Verbs *abgeben* die Nutzung von *Gingko* in *KorAP* illustrieren.

Gingko ist für die Vermittlung der ingenieurwissenschaftlichen Fachsprache hochrelevant. Hintergrund ist, dass vielerorts mehr sprachlicher Unterricht für internationale Studierende (vgl. Wisniewski et al. 2022) und auch speziell für Studierende der Ingenieurwissenschaften (vgl. Heine / Oehmichen / Roß 2019) gefordert wird. Zudem sind die Studierenden dieser Fächergruppe mit 42 % aller internationalen Studierenden an deutschen Hochschulen die größte Zielgruppe unter den Lernenden des Deutschen als fremder Wissenschaftssprache (DAAD / DZHW 2022: 5). Ein ausreichend großes Korpus ist in diesem Kontext eine wichtige Grundlage, um Unterrichtsinhalte empirisch zu legitimieren. So können etwa ermittelte Häufigkeiten sprachlicher Phänomene didaktische Entscheidungen zur Priorisierung von Unterrichtsgegenständen begründen (vgl. Fandrych / Tschirner 2007: 197). *Gingko* ist das erste öffentliche Korpus der technischen (fach-)wissenschaftlichen Sprache. Auch die öffentlichen Korpora anderer wissenschaftlicher Disziplinen sind überschaubar: Die als wissenschaftlich gekennzeichneten Subkorpora der DWDS-Kernkorpora beinhalten v.a. Lexikonartikel und geisteswissenschaftliche Texte; die DWDS-Webkorpora zu Jura und Medizin, das *Fachsprachenkorpus* und *spektrumdirekt* aus DeReKo enthalten stärker populärwissenschaftliche Texte.

Gingko wurde im DFG-geförderten Projekt „Muster in der Sprache der Ingenieurwissenschaften – *Gingko* (Geschriebenes ingenieurwissenschaftliches Korpus)“ (2017-2022; Kennzeichen: HE 7157/1-

1; www.uni-leipzig.de/gingko) an der Universität Greifswald (Leitung: Jun. Prof. Antje Heine) und dem Herder-Institut der Universität Leipzig (Leitung: Prof. Christian Fandrych) und in Zusammenarbeit mit dem Karlsruher Institut für Technologie und dem IDS erstellt. Das Korpus besteht aus Artikeln der Zeitschriften *Automobiltechnische Zeitschrift* (ATZ) und *Motortechnische Zeitschrift* (MTZ) des Springer-Verlags (Jahrgänge 2007 bis 2016). Der Adressatenkreis sind Institutionen aus Forschung und Entwicklung, Unternehmen, Hersteller oder Zulieferer. Ein großer Teil der Beiträge beruht auf Dissertationen und Forschungsprojekten. Das Korpus umfasst 2498 Texte und 4.667.656 Millionen Token¹. Die durchschnittliche Textlänge beträgt 2000 Token. Die Metadaten beinhalten Informationen zur Veröffentlichung der Artikel, wie Autor:in(nen), Titel oder DOI. Für Lemmata, Wortarten (*Part-of-Speech*) und Satzspannen stehen Annotationen unterschiedlicher Tagger zur Verfügung. Weiterhin sind syntaktische Abhängigkeiten und Konstituenten sowie die Dokumentstruktur, also beispielsweise Überschriften oder Absätze, annotiert. Genauere Informationen zu allen Metadaten, Annotationen und zur Korpusabfrage können der KorAP-Einführung², der Online-Hilfe zu *COSMAS II*³ oder dem Korpushandbuch (vgl. Portmann et al. 2022) entnommen werden.

Im Rahmen des Gingko-Projekts wurden vielfältige Analysen zur sprachlichen Gestaltung der Sprache der Ingenieurwissenschaften durchgeführt. Im Fokus standen dabei neben unpersönlichen Ausdrucksmustern (vgl. Heine / Schirrmeister / Anderson 2019), Nominalisierungen (vgl. Heine / Rummel 2019; Rummel / Heine 2019) und Attributrealisierungen (vgl. Meißner 2022) die figurativen Verben (vgl. Heine 2019). Figurative Verben gelten als charakteristisch für die alltägliche Wissenschaftssprache und stellen eine Schwierigkeit für DaF-Lernende dar, da sie in der Wissenschaftssprache z.T. andere Bedeutungen als in der Alltagssprache tragen (vgl. Ehlich 1995; Fandrych 2006; Meißner 2014: 51-55). Heine (2019) fragt nach Unterschieden in der Verteilung figurativer Verben in verschiedenen wissenschaftlichen Disziplinen. Sie stellt fest, dass ein großer Anteil der als figurativ angenommenen Verben in geisteswissenschaftlichen Texten häufiger vorkommt als in ingenieurwissenschaftlichen Texten. Sie räumt jedoch ein, dass die rein formbasierte Analyse nicht ausreicht, um etwas über die Relevanz von figurativen Lesarten der jeweiligen Verben in den unterschiedlichen Disziplinen auszusagen. Sie zeigt für das Beispiel von *eintreten*, dass sich allein in den Belegen in *Gingko* unterschiedliche Lesarten mit verschiedenen Graden von Figurativität ausmachen lassen und formuliert das Desiderat, nicht nur Wortformen, sondern auch Lesarten zwischen Korpora zu vergleichen.

Zur Illustration der Nutzung von *Gingko* in KorAP soll im Folgenden analog zum Fallbeispiel von Heine (2019) eine semantische Analyse durchgeführt werden, die zusätzlich ermitteln soll, ob sich die Verwendung der verschiedenen Lesarten von Partikelverben in wissenschaftlicher und journalistischer Sprache unterscheidet. Dafür wird neben *Gingko* das Wochenmagazin FOCUS als ein anderes Subkorpus von DeReKo analysiert. Als ein beliebiges, beispielhaftes Partikelverb wird *abgeben* ausgewählt.

Zunächst wird über die Metadatenvariable *corpusSigle* die Suche auf die Zeitschriften ATZ und MTZ begrenzt. Die einfache Lemmaabfrage in der *ANNIS Query Language* (vgl. Krause / Zeldes 2016) lautet *l="abgeben"*. Abbildung 1 zeigt den Screenshot der zugehörigen Ergebnisansicht, die Belege (1) und (2) zwei Treffer aus dem Screenshot mit mehr Kontext. Um Verb und Partikel in Distanzstellung zu

¹ Die Forschungsarbeiten von Heine / Schirrmeister / Anderson (2019), Heine / Rummel (2019), Rummel / Heine (2019) und Meißner (2022) und nutzen eine Version des Korpus vor der Veröffentlichung in DeReKo. In dieser Version 1.0 sind es noch 2517 Texte und 4.693.821 Token.

² <https://korap.ids-mannheim.de/> (12.04.2022).

³ <https://www2.ids-mannheim.de/cosmas2/web-app/hilfe/suchanfrage/> (12.04.2022).

finden, kann nach *ab* und *geben* gesucht werden, die durch die Relation „abgetrennter Verbzusatz“ verbunden sind⁴: `l="/ab/->malt/d[func="AVZ"] l="/geben/`. Die Belege (3) und (4) sind Treffer dieser Suchanfrage.



Abbildung 1
Ergebnisansicht in *Gingko*

- (1) Weil Elektromotoren jedoch mit steigender Temperatur weniger Leistung **abgeben**, [...] (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 Leakage, DOI: 10.1007/BF03225612)
- (2) Bevor die an Sauerstoff abgereicherte Abluft an die Umgebung **abgegeben** wird, [...] (ATZ16/OKT/00001, Automobiltechnische Zeitschrift, Oktober 2016, Nr. 118(10), S. 28-37, Pischinger, S.; Ogrzewalla, J.: Integration eines Brennstoffzellen-Range-Extenders, DOI: 10.1007/s35148-016-0099-9)
- (3) Laut Behr **geben** Batterien nur 5 % als Abwärme **ab**. (ATZ11/NOV/00001, Automobiltechnische Zeitschrift, November 2011, Nr. 113(11), S. 810-811, Thermomanagement: Neue Perspektiven für das E-Auto, DOI: 10.1365/s35148-011-0186-x)
- (4) Weitere Erklärungen für das gute Abschneiden im Vergleich zu deutschen TUs **gab** die Medienstelle nicht **ab**. (MTZ16/APR/00011, Motortechnische Zeitschrift, April 2016, Nr. 77(4), S. 8-13, Hofacker, A.: Internationale Rankings Der unsichtbare Erfolg deutscher TUs, DOI: 10.1007/s35146-016-0034-3)

⁴ Die Relation ist eine Annotation des *MaltParsers* (vgl. Nivre / Hall / Nilsson 2006). Stattdessen können für die Suche auch die unterschiedlichen Versionen des POS-Taggings genutzt werden. Man kann nach den Lemmata *geben* und *ab* suchen, die als finites Vollverb bzw. als Partikel getaggt sind und innerhalb eines Satzes stehen:

Poliqarp+: `matches(contains(<base/s=s>, [l="ab" & p="PTKVZ"])), contains (<base/s=s>, [l=geben]))`
ANNIS: `base/s="s" & l="geben" & l="ab" & p="PTKVZ" & #1 _i_ #2 & #1 _i_ #3 & #3 _= #4 & #2 . * #3`

In KorAP lässt sich das Suchergebnis als csv-Datei exportieren, sodass die Lesarten der einzelnen Belege in einem Tabellenkalkulationsprogramm manuell nachannotiert werden können. Aus den mit und ohne Distanzstellung vorkommenden Strukturen in *Gingko* wurden insgesamt 30 Belege zufällig ausgewählt. 27 Belege konnten der Bedeutung „etwas sondert etwas ab und überträgt es auf etwas“⁵ zugeordnet werden – wie auch Beleg (1) bis (3). Die anderen drei Belege sind Funktionsverbgefüge, die einen kommunikativen Akt beschreiben (vgl. 4): *eine Erklärung/Empfehlung/Einschätzung abgeben*. Für dieselbe Suche von *abgeben* im Wochenmagazin FOCUS (corpusSigle eq /FOC../) und nach einer Annotation von 30 Belegen wurden nur zwei Belege mit der erstgenannten Bedeutung gefunden, dagegen waren es 12 Funktionsverbgefüge (vgl. 5) sowie neun Belege mit der Lesart „jemand übt etwas nicht mehr aus und überlässt es jemandem; abtreten“⁶ (vgl. 6). Nur zwei Belege hatten in FOCUS die Bedeutung „etwas sondert etwas ab und überträgt es auf etwas“.

- (5) [...] Empfehlung **geben** die Sachverständigen zwar bewusst nicht **ab**, [...] (FOC21/OKT/00024, FOCUS, 02.10.2021, S. 66. Originalressort: Konjunktur, Interview: Carla Neuhaus, Große Steuersenkungen sind nicht drin)
- (6) [...], dass die Landtage ihre Macht an 100 Beamte **abgeben** [...] (FOC02/MAI/00191, FOCUS, 13.05.2002, S. 052-052, Originalressort: Deutschland: Peter Glotz, Der Föderalismus in Deutschland hat versagt)

In den journalistischen Texten scheinen dieser ersten Analyse von 30 Belegen zufolge also mehr Funktionsverbgefüge als in den ingenieurwissenschaftlichen Texten vorzukommen. Die Bedeutung des Funktionsverbs *abgeben* könnte am Ende einer angenommenen Abstraktionsskala eingeordnet werden: In den gefundenen Funktionsverbgefügen drückt das Funktionsverb *abgeben* nur noch die Abgeschlossenheit der Kommunikation aus. Als konkrete Grundbedeutung am Anfang der Skala könnte die Lesart „jemandem etwas aushändigen“⁷ gegenübergestellt werden. Die in *Gingko* vorherrschende Bedeutung „etwas sondert etwas ab und überträgt es auf etwas“ stellt einen kleineren Schritt hin zur Bildhaftigkeit dar: Nicht Personen sind in eine Handlung involviert, sondern Materialien in einen Vorgang.

Das Ergebnis veranschaulicht zunächst, dass Studierende in beiden Varietäten auf Figurativität und übertragene Bedeutungen stoßen. Studienanfänger:innen treffen in ingenieurwissenschaftlichen Texten jedoch wahrscheinlich auf Lesarten, denen sie in studienvorbereitenden Lehrmaterialien zumindest weniger häufig begegnen. In sprachlichen Hochschulzulassungstests, auf die diese Kurse vorbereiten, wird i.d.R. auf populärwissenschaftlich oder journalistisch geprägte Texte zurückgegriffen, da im Leseverstehen ein „wissenschaftsorientierter Text“ (HRK / KMK 2004/2019: 12) gefordert wird, der keine Fachkenntnisse voraussetzen soll. In universitären Sprachkursen könnten beispielhafte Korpusbelege wie die hier präsentierten Lernende auf Stufungen der Bildhaftigkeit aufmerksam machen und für Unterschiede zwischen Alltagssprache und Wissenschaftssprache sensibilisieren.

Mit der Integrierung von *Gingko* in DeReKo steht das Korpus für weitere Analysen kostenlos zur Verfügung. Zusätzlich arbeitet das IDS kontinuierlich an den Funktionalitäten von KorAP und auch des RKorAPClients (vgl. Kupietz / Diewald / Margaretha 2020), der an dieser Stelle nur erwähnt werden kann. Wünschenswert wäre, dass dem Korpus in Zukunft regelmäßig aktuelle Texte hinzugefügt werden können, da sich wegen rascher Entwicklungen gerade in der Automobiltechnik auch die Sprache der Ingenieurwissenschaften verändert.

⁵ Elektronisches Valenzwörterbuch deutscher Verben. Permalink: <https://grammis.ids-mannheim.de/verbs/view/400206/> (12.04.2022).

⁶ <https://grammis.ids-mannheim.de/verbs/view/400206/> (12.04.2022).

⁷ <https://grammis.ids-mannheim.de/verbs/view/400206/> (12.04.2022).

Literatur und Ressourcen

COSMAS II: Leibniz-Institut für Deutsche Sprache (IDS). <https://cosmas2.ids-mannheim.de> (12.04.2022).

DAAD / DZHW (2022): *Wissenschaft weltoffen kompakt 2022. Daten und Fakten zur Internationalität von Studium und Forschung in Deutschland*. wbv.

DWDS: Berlin-Brandenburgische Akademie der Wissenschaften (BBAW). <https://www.dwds.de> (12.04.2022).

Ehlich, Konrad (1995): Die Lehre der deutschen Wissenschaftssprache. Sprachliche Strukturen, didaktische Deseiderate. In: Kretzenbacher, Heinz Leonhard (Hrsg.): *Linguistik der Wissenschaftssprache*. Berlin: De Gruyter, 325-351.

Elektronisches Valenzwörterbuch deutscher Verben: IDS. <https://grammis.ids-mannheim.de/verbvalenz> (12.04.2022).

Fandrych, Christian (2006): Bildhaftigkeit und Formelhaftigkeit in der allgemeinen Wissenschaftssprache als Herausforderung für Deutsch als Fremdsprache. In: Ehlich, Konrad / Heller, Dorothee (Hrsg.): *Die Wissenschaft und ihre Sprachen*. Frankfurt: Peter Lang, 39-62.

Fandrych, Christian / Tschirner, Erwin (2007): Korpuslinguistik und Deutsch als Fremdsprache. Ein Perspektivenwechsel. In: *Deutsch als Fremdsprache* 4, 195-204.

Heine, Antje (2019): Wozu (ge-)brauchen Ingenieure figurative Verben? Eine Analyse auf Basis wissenschaftlicher Artikel. In: *Informationen Deutsch als Fremdsprache* 47: 6, 612-629. <https://doi.org/10.1515/infodaf-2019-0057> (12.04.2022).

Heine, Antje / Oehmichen, Kathleen / Roß, Tilo (2019): *Zur Sprachkompetenz internationaler Studierender. Befragung von Studierenden und Lehrenden in den Ingenieurwissenschaften*. <https://tu-dresden.de/bu/verkehr/iad/lvm/studium/sprachkompetenz> (12.04.2022).

Heine, Antje / Rummel, Marlene (2019): Funktion statt Transformation. Kausale Relationen in wissenschaftlichen Texten. In: Busch-Lauer, Ines-Andrea (Hrsg.): *DaFF-Impulse – Deutsch als Fremd- und Fachsprache an Hochschulen*. Berlin: Frank & Timme, 11-32.

Heine, Antje / Schirrmeyer, Lars / Anderson, Matthew O. (2019): The Better the Description, the Better the Instruction. Passives and Impersonals in the German Language of (Automotive) Engineering. In: *German as a foreign language* 2. <http://www.gfl-journal.de/2-2019/heine.pdf> (12.04.2022).

HRK / KMK (2004|2019): *Rahmenordnung über Deutsche Sprachprüfungen für das Studium an deutschen Hochschulen (RO-DT). Beschluss der HRK vom 08.06.2004 und der KMK vom 25.06.2004 i.d.F. der HRK vom 23.07.2020 und der KMK vom 28.11.2019*. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_06_25_RO_DT.pdf (12.04.2022).

Krause, Thomas / Zeldes, Amir (2016): ANNIS3: A new architecture for generic corpus query and visualization. In: *Digital Scholarship in the Humanities* 31: 1, 118-139. <https://doi.org/10.1093/llc/fqw057> (12.04.2022).

Kupietz, Marc / Diewald, Nils / Margaretha, Eliza (2020): RKorAPClient: An R Package for Accessing the German Reference Corpus DeReKo via KorAP. In: Calzolari, Nicoletta et al. (Hrsg.): *Proceedings of The 12th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, 7015-7021.

Kupietz, Marc et al. (2020): Recherche in Social-Media-Korpora mit KorAP. In: Marx, Konstanze / Lobin, Henning / Schmidt, Axel (Hrsg.): *Deutsch in Sozialen Medien*. Berlin: De Gruyter, 373-378.

Kupietz, Marc et al. (2018): The German Reference Corpus DeReKo: New Developments – New Opportunities. In: Calzolari, Nicoletta et al. (Hrsg.): *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC'18)*. Miyazaki: European Language Resources Association (ELRA), 4353-4360.

Meißner, Cordula (2014): *Figurative Verben in der allgemeinen Wissenschaftssprache des Deutschen. Eine Korpusstudie*. Tübingen: Stauffenburg.

Meißner, Cordula (2022): Attributrealisierung und Nominalstil in der Sprache der Ingenieurwissenschaften. Kollexemanalysen zum Gebrauch von Partizipialattributen und Relativsätzen. In: *Deutsch als Fremdsprache* 2, 88-104.

Nivre, Joakim / Hall, Johan / Nilsson, Jens (2006): MaltParser: A Data-Driven Parser-Generator for Dependency Parsing. In: Calzolari, Nicoletta et al. (Hrsg.): *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. Genua: European Language Resources Association (ELRA).

Portmann, Annette et al. (2022): *Gingko (Geschriebenes ingenieurwissenschaftliches Korpus). Handbuch*.

Rummel, Marlene / Heine, Antje (2019): Von lexikalischen Markern zu komplexen Mustern – zum semantischen Profil automobiltechnischer Fachtexte. In: *Fachsprache* 41: 3-4, 123-141. <https://doi.org/10.24989/fs.v41i3-4.1726> (12.04.2022).

Schirrmeister, Lars et al. (2021): Gingko – ein Korpus der ingenieurwissenschaftlichen Sprache. In: *Deutsch als Fremdsprache* 4, 214-224. <https://doi.org/10.37307/j.2198-2430.2021.04.04> (12.04.2022).

Wisniewski, Katrin et al. (Hrsg.) (2022): *Sprache und Studienerfolg bei Bildungsausländer/-innen*. Münster: Waxmann.

Biographische Notiz: Annette Portmann ist wissenschaftliche Mitarbeiterin im Forschungsprojekt „Muster in der Sprache der Ingenieurwissenschaften: Gingko - Geschriebenes ingenieurwissenschaftliches Korpus“ am Herder-Institut der Universität Leipzig. Ihre Forschungsinteressen liegen im Bereich Deutsch als fremde Wissenschaftssprache, Lernaltersprache und Korpuslinguistik.

Kontaktanschrift:

Annette Portmann

Herder-Institut

Universität Leipzig

annette.portmann@uni-leipzig.de

