



La revue *Aleph. langues, médias et sociétés* est approuvée par ERIHPLUS. Elle est classée à la catégorie B.


Korpora als praktische Werkzeuge wissenschaftlicher Analyse. Am Beispiel vom Referenzkorporus DeReKo des Leibniz-Instituts der Deutschen Sprache

المجموعات المرجعية كأداة عملية في التحليل العلمي. مساهمة مبنية على المجموعة DeReKo المرجعية للمعهد لايبنتز للغة الأمانيد

Les corpus comme outils pratiques d'analyse scientifique. Cas de DeReKo de Leibniz-Institut de la Langue Allemande

Corpora as practical tools for scientific analysis. Using the example of the DeReKo reference corpus of the Leibniz-Institute of the German Language

Osmane Baya - Alger 2

	Soumission	Publication numérique	Publication Asjp
	12-11- 2023	14-01-2024	15-01-2024

Éditeur : Edile (Edition et diffusion de l'écrit scientifique)

Dépôt légal : 6109-2014

Edition numérique : <https://aleph.edinum.org>

Date de publication : 14 janvier 2024

ISSN : 2437-1076

(Edition ASJP) : <https://www.asjp.cerist.dz/en/PresentationRevue/226>

Date de publication : 15 janvier 2024

Pagination : 53-65

ISSN : 2437-0274

Référence papier

Baya Osmane, « Korpora als praktische Werkzeuge wissenschaftlicher Analyse. Am Beispiel vom Referenzkorporus DeReKo des Leibniz-Instituts der Deutschen Sprache », *Aleph*, Vol 11 (1) | 2024, 53-65.

Référence électronique

Baya Osmane, « Korpora als praktische Werkzeuge wissenschaftlicher Analyse. Am Beispiel vom Referenzkorporus DeReKo des Leibniz-Instituts der Deutschen Sprache », *Aleph* [En ligne], Vol 11 (1) | 2024, mis en ligne le 14 janvier 2024.. URL : <https://aleph.edinum.org/10547>

Korpora als praktische Werkzeuge wissenschaftlicher Analyse. Am Beispiel vom Referenzkorpus DeReKo des Leibniz-Instituts der Deutschen Sprache

المجموعات المرجعية كأداة عملية في التحليل العلمي. مساهمة مبنية على المجموعة DeReKo
المرجعية للمعهد لايبنيز للغة الألمانية

**Les corpus comme outils pratiques d'analyse scientifique.
Cas de DeReKo de Leibniz-Institut de la Langue Allemande**

**Corpora as practical tools for scientific analysis. Using the
example of the DeReKo reference corpus of the Leibniz-
Institute of the German Language**

BAYA OSMANE

الجزائر 2

Einleitung

Seit geraumer Zeit besteht eine hohe Nachfrage nach zuverlässigen Hilfsmitteln zur Erstellung wissenschaftlicher Analysen, wobei Korpora eine immer größere Grundlage bilden. Korpora sind aufbereitete Textsammlungen, die in unterschiedlichen Bereichen verwendet werden können und einen ganz konkreten Referenzbereich abdecken. Sie können zur Erforschung sprachlicher Erscheinungen, zur Erstellung von Grammatiken, im Sprachenvergleich oder im Fremdsprachenunterricht, aber auch zur Beantwortung bestimmter sprachwissenschaftlicher Untersuchungsfragen genutzt werden. Natürlich sollten beim Erstellen eines Korpus bestimmte Leitfragen gestellt werden, die zur Klärung der Forschungsfrage beitragen.

Das wissenschaftliche Programm der Korpuslinguistik am Leibniz-Institut für Deutsche Sprache ist eine moderne Methode innerhalb der Linguistik. Das sehen auch McEnery und Wilson so: “[...] this modern phenomenon, corpus linguistics, has come to be an increasingly prevalent methodology in linguistics [...]” (McEnery/Wilson, 2003: 1). Das Institut beheimatet aktuell die größte Korpusammlung in deutscher Sprache. Sie ist entstanden durch die explorative Analyse von sehr großen Sammlungen natürlichsprachlicher Daten, die neue Einblicke in die Strukturen, Gesetzmäßigkeiten, Eigenschaften und Funktionen der deutschen Sprache geben.

In diesem Beitrag wird der Versuch unternommen, die praktische Bedeutung eines Korpus bzw. der Korpuslinguistik für sprachwissenschaftliche Analysen zu beleuchten. Darüber hinaus wird die Aufmerksamkeit auf die

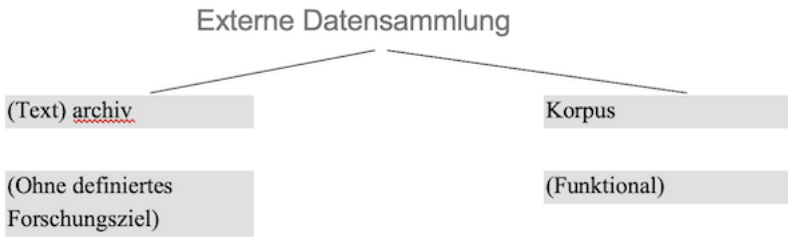
Korpuslandschaft DeReKo, die Ressourcen und Analysewerkzeuge, gerichtet und in den Mittelpunkt gerückt.

1. Was ist ein Korpus?

Ein Korpus ist eine Sammlung aufbereiteter Texte. “Ein Korpus ist eine Sammlung schriftlicher oder gesprochener Äußerungen in einer oder mehreren Sprachen in digitaler Form.” (Lemnitzer/Zinsmeister, 2006: 13). Es dient als Arbeitsgrundlage für Analysen, um bestimmte Forschungsfragen zu beantworten. Das Ziel der Erstellung von Korpora ist unter anderem, neue Erkenntnisse über Strukturen, Funktionen und die Verwendung natürlicher Sprachen in schriftlicher Form zu erlangen. Durch den Einsatz großer annotierter Korpora kann man zu statistisch signifikanten Ergebnissen kommen, die eventuell in kleineren Untersuchungen nicht erkennbar wären.

Bußmann definiert “Korpus” im Lexikon der Sprachwissenschaft folgendermaßen: Corpus [Neutr., Pl. Corpora; lat. corpus ‘Körper’]. Endliche Menge konkreter sprachlicher Äußerungen, die als empirische Grundlage für sprachwissenschaftliche Untersuchungen dienen. Der Stellenwert und die Beschaffenheit des Korpus hängen weitgehend von den spezifischen Fragestellungen und methodischen Voraussetzungen des theoretischen Rahmens der Untersuchung ab, wie sich z.B. an der unterschiedlichen Einschätzung empirischer Daten im Strukturalismus und in der Generativen Syntax zeigt: Während der Strukturalismus bei der Beschreibung sprachlicher Strukturen ausschließlich von beobachtbaren Corpora sprachlicher Äußerungen ausgeht, sich induktiver Aufdeckungsprozeduren (Segmentierung) bedient, die Intuition der Forschenden als Beurteilungsinstanz ablehnt und die Gültigkeit der Aussagen auf das jeweils zugrunde liegende Korpus einschränkt, spielen Corpora in der generativen Grammatik keine wesentliche Rolle (Bussmann, 2008: 143).

In diesem Beitrag wird zwischen Texten unterschieden, die digitalisiert werden, um sie möglicherweise später auszuwerten (im Sinne von Textarchiven), und Texten, die bereits ein definiertes Forschungsziel haben und aufbereitet sind. Was ein Korpus von einer Sammlung authentischer Daten (Textarchive) unterscheidet, die schon seit langer Zeit verwendet werden, ist die “digitale Form” sowie die Korpusaufbereitung und die Verwendung von “korpuslinguistischen Tools” anstelle manueller Verfahren für ihre Analyse. Das folgende Schema zeigt die Unterscheidung zwischen externen Datensammlungen in Textarchiven und Korpora: (Vgl. Wegera/Kwekkeboom/Herbers/Schultz-Balluff, 2023: 38)



Schema 1: Aufteilung der externen Datensammlung

Das in Korpora aufgezeichnete Ergebnis wird als Grundlage sowohl für die explorative als auch induktive Analyse, auf Theoriebildung zielende Generalisierungsstrategie verstanden.

2. Nach welchen Kriterien wird ein virtuelles Korpus erstellt?

Bei der Erstellung eines Korpus sollten wichtige grundlegende Fragen geklärt werden. Carmen Scherer hat dabei die folgenden Fragen betont: Die Frage nach dem Zweck, die ein Korpus erst zu einem Korpus macht und ihn von anderen Textsammlungen abgrenzt, die Frage nach der Repräsentativität, der Beständigkeit, dem Inhalt sowie der Korpusgröße. (Vgl. Scherer, 2006: 5-10) Es sollten zum Beispiel zeitliche, räumliche oder textsortenspezifische Einschränkungen getroffen werden, um den Zweck des Korpus genauer zu bestimmen. Diese Beschränkungen können aber erst nach der Erstellung der Forschungsfrage entschieden werden. Natürlich sollten Korpora, damit sie verlässlich durchsuchbar und im Internet verfügbar sind, in adäquaten Dateiformaten aufbereitet werden. In diesem Abschnitt soll ein Überblick über die wichtigsten Kriterien zur Erstellung eines virtuellen Korpus gegeben werden.

Als erstes Kriterium gilt der Zweck des Korpus. Er ist ein sehr wichtiges Kriterium, da er auch einen Einfluss auf die weiteren Kriterien der Repräsentativität, der Annotation und der Korpusgröße hat. Man sollte die Funktionalität in den Vordergrund stellen, um die relevantesten Texte zu selektieren. Die Korpus-Funktionalität kann im Laufe der Zeit variieren. Durch diese Funktionalität entscheidet man sich für die Sprache des Korpus und für das Medium, das sich auf die Primärdaten bezieht. Es kann also ein Korpus sein, das mono-, bi- oder multilingual sein kann; außerdem unterscheidet man zwischen Korpora der geschriebenen oder gesprochenen Sprache (Vgl. Flinz, 2021).

Das zweite Kriterium ist der Bezug zum Untersuchungsgegenstand. Er ist mit den Eigenschaften des zu repräsentierenden Gegenstandes verbunden, da sie in ihrer Auswahl repräsentativ für den Referenzbereich und die gewählte Grundgesamtheit sein müssen. Die Frage der Repräsentativität sollte trotzdem dem Benutzer überlassen sein, in Abhängigkeit von der jeweiligen Forschungsfrage (Vgl. Flinz, 2021: 7).

Das dritte Kriterium ist die Annotation. Bei der Annotation werden die Daten mit linguistischen und/oder semantischen Informationen angereichert und mit Anmerkungen versehen. (Vgl. Flinz, 2021: 8) Diese Informationen können Part-of-Speech-Markierungen (Zuordnung von Wörtern zu Wortarten), Lemmata (Stichworte) und Entitäten (Personen, Organisationen, Unternehmen etc.) umfassen. Die Wahl des Annotationsschemas hängt von der Forschungsfrage und der Art der Forschungsanalyse ab.

Das vierte Kriterium ist die Größe eines Korpus. Der Umfang von Textdaten spielt eine wichtige Rolle bei quantitativen Untersuchungen, da valide statistische Aussagen nur möglich sind, wenn eine große Menge an Daten vorhanden ist. «Eine Maximierung der Korpusgröße ist aus korpuslinguistischer Perspektive grundsätzlich wünschenswert, denn je größer das Korpus, desto verlässlichere Aussagen über mehr verschiedene und seltenere Phänomene können anhand des Korpus getroffen werden» (Lüngen, 2017: 164). Trotzdem kann die Größe stark variieren. Wenn man beispielsweise Spezialkorpora untersucht, kann auch eine begrenzte Anzahl an Textwörtern ausreichen (Vgl. Flinz/Perkuhn, 2018; Flinz 2021). Die Größe eines Korpus ist durchaus mit der Forschungsfrage verbunden.

3. Wie erstellt man ein eigenes Korpus für wissenschaftliche Analysen?

Für viele Forschungsfragen können bereits bestehende öffentlich verfügbare Korpora verwendet werden. Doch oft ist es trotzdem sinnvoll, ein eigenes Korpus für ein Forschungsvorhaben zusammenzustellen, das den eigenen Bedürfnissen möglichst gut entspricht. Bei der Durchführung eines eigenen Projekts sind folgende Schritte hilfreich:

- Die Forschungsfrage definieren: Bei der Erstellung eines Korpus sollte zuerst die Forschungsfrage definiert werden. Sie bildet den Schwerpunkt einer wissenschaftlichen Untersuchung und gibt die Rahmenbedingungen für die Analyse vor. Die Forschungsfrage sollte als Leitfaden für die Auswahl der Texte und die Wahl der Ergänzungen dienen.

- Texte auswählen: Es stehen zahlreiche vorgefertigte Korpora online zur Verfügung. Texte können nach synchronen und/oder diachronen Kriterien ausgewählt werden. Dennoch ist es wichtig zu beachten, ob die ausgewählten Texte aus urheberrechtlichen Gründen für die Untersuchung verwendet werden dürfen. Bei einigen Korpora ist die Verfügbarkeit an eine Registrierung oder an einen Lizenzvertrag gebunden.
- Das Korpus analysieren und organisieren: Nach der Zusammenstellung des Korpus erfolgt die eigentliche Analyse. Dabei wird der Forschungsfrage nachgegangen und es wird versucht, potenzielle Antworten mit dem Korpus zu liefern. Oft liegt das Interesse auf dem Gebrauch eines Wortes im Korpus. Man kann eine spezielle Suchanfrage starten, zum Beispiel wie im DeReKo (siehe Abb. 4). Der letzte Schritt besteht darin, das Korpus zu organisieren. Die Informationsstrukturierung ist wichtig, um festzulegen, welche Informationen mithilfe des fertigen Korpus abgerufen werden sollen. Je nach Forschungsfrage und den spezifischen Zielen der Analyse kann das Korpus nach Themen geordnet werden.

4. Zum Programm der Korpuslinguistik am Leibniz-Institut für Deutsche Sprache (IDS) in Mannheim

Für wissenschaftliche Forschung und die Bearbeitung linguistischer Forschungsfragen sind Textdaten unerlässlich. Die Korpuslinguistik beschäftigt sich mit dem Aufbau, der Aufbereitung und der Auswertung von elektronischen Korpora. Korpuslinguistik ist eine wissenschaftliche Tätigkeit und muss wissenschaftlichen Prinzipien folgen und wissenschaftlichen Ansprüchen genügen (Vgl. Lemnitzer/Zinsmeier, 2006: 9).

Das wissenschaftliche Programm der Korpuslinguistik am IDS wird von der explorativen Analyse sehr großer Sammlungen natürlichsprachlicher Daten bestimmt, um neue Erkenntnisse über die Strukturen, Gesetzmäßigkeiten, Eigenschaften und Funktionen von Sprache zu gewinnen. Das Leibniz-Institut hat sich folgende Ziele für das Programm der Korpuslinguistik gesetzt:

- Der Programmbereich ist am IDS dafür verantwortlich, den deutschen Schriftsprachgebrauch beständig und in angemessener Weise stichprobenartig im Deutschen Referenzkorpus zu dokumentieren.
- Ausgehend von grundsätzlichen Überlegungen zur linguistischen Theoriebildung wird eine Methodik der Korpusanalyse und

-erschließung entwickelt, die auf mathematisch-statistischen Methoden basiert.

- Die gewonnenen Erkenntnisse werden auf wissenschaftstheoretischer Ebene reflektiert und in die Diskussion der linguistischen Theoriebildung eingebracht.
- Darüber hinaus wird die erarbeitete Methodik in Kooperationen mit anderen, teils externen Projekten zur Gewinnung und linguistischen Beschreibung üblicher Wortverbindungen und anderer präferenzrelationaler Strukturen eingebracht.

4.1. Das Deutsche Referenzkorpus (DeReKo)

DeReKo (Deutsches Referenzkorpus) sind Korpora, die am Leibniz-Institut erarbeitet wurden. Sie bilden mit 55 Milliarden Wörtern (Stand 08.03.2023) die derzeit größte linguistisch motivierte Sammlung von Korpora mit geschriebenen deutschsprachigen Texten aus der Gegenwart und der neueren Vergangenheit. Das Korpus wird fortlaufend erweitert und das bestehende Korpusmaterial überarbeitet. Das DeReKo ist in mehrere Unterkorpora aufgeteilt, die teilweise auch morphologisch und syntaktisch annotiert sind. Unter den enthaltenen Korpora befinden sich belletristische, wissenschaftliche und populärwissenschaftliche Texte, aber auch historische Korpora (u.a. Texte von Marx und Engels, Goethe und den Brüdern Grimm), diverse Regionalzeitungen, die sehr viele Regionen Deutschlands abdecken (Mannheimer Morgen und Bonner Zeitungskorpus, Braunschweiger Zeitung), Zeitschriften (z.B. Focus) und Zeitungen (ZEIT, taz, SZ), Wikipedia-Artikel und Diskussionen, das Wendekorpus (Sprache in West- und Ostdeutschland 1989/90), literarische Texte (u.a. von Christa Wolf und Martin Walser) und Sonderkorpora etwa zu Fachsprachen, CMC-Korpora (Dortmunder Chatkorpus) und vergleichbare Korpora wie (die Wikipedia Korpora). Neu in DeReKo-2022 ist das Gingko-Korpus des Projektes «Muster in der Sprache der Ingenieurwissenschaften». Zudem sind auch internationale deutschsprachige Korpora aus Österreich, Luxemburg und der Schweiz vorhanden.

Aus urheberrechtlichen Gründen ist nur ein Teil des IDS-Korpus öffentlich zugänglich, und auch dieser Teil ist erst nach kostenloser Registrierung für nicht-kommerzielle Zwecke verfügbar. DeReKo lässt sich – wie bereits erwähnt – nach kostenloser Registrierung mit dem COSMAS-System durchsuchen, insbesondere mit der Web-Anwendung COSMAS II web.

Eine Übersicht über das Archiv finden Sie <https://colibris.link/HMevX> und <https://colibris.link/rfV6e>

Beim Ausbau von DeReKo wurde auf Diversität geachtet. Selbstverständlich spielen auch langfristige Akquirierungen eine Rolle. Der Korpusaufbau bemüht sich mit Akquisitionskampagnen um die Verwendungsrechte der Verlage. Er bemüht sich allerdings auch um Angebote (von Rechte-Inhabern) und Anfragen (von IDS-internen und -externen Projekten und Nutzern) sowie um Kosten/Nutzen-Abwägungen bezüglich des finanziellen und technischen Aufwands der Akquisition.

4.2. Die wesentlichen Schritte zur Benutzung von DeReKo

Der Zugriff auf DeReKo erfolgt über das webbasierte Korpusrecherchesystem COSMAS II (Corpus Search, Management and Analysis System). COSMAS II ermöglicht es dem Benutzer, DeReKo nach linguistischen Kriterien zu durchsuchen. Es erlaubt die Bildung unterschiedlicher Suchanfragen, einschließlich Wörtern, Teilwörtern, Grundformen, Wortklassen, grammatikalischen Mustern usw. Zudem ermöglicht es die Darstellung der Ergebnisse nach individuellen Kriterien sowie die Dokumentation und den Export der zugehörigen Belege (Vgl. Bopp, 2010).

Als erster Schritt ist es erforderlich, sich in COSMAS II zu registrieren. Die Registrierung erfolgt unter <https://perso.ids-mannheim.de/registration/> und anschließend sollten Sie sich einloggen. Die Registrierung dauert nur wenige Minuten. Unmittelbar danach erhalten Sie eine Bestätigungsmail mit Ihrem Benutzernamen und Ihrem Passwort. Die folgende Abbildung zeigt, wo Sie sich registrieren und einloggen können.

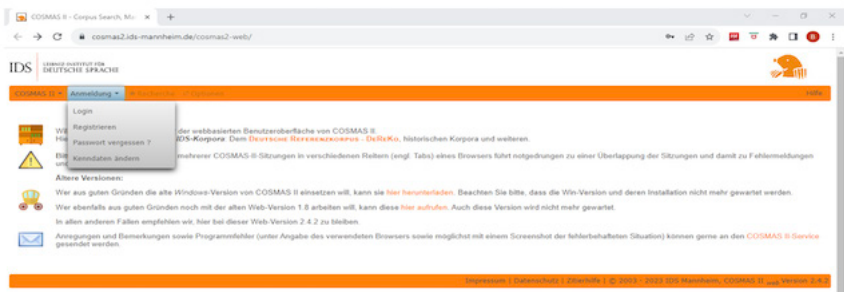


Abb. N° 1 : Registrierung und Einloggen

Als nächster Schritt erscheinen zahlreiche verschiedene Teilkorpora, die meist durch u. a. historische, Literatur-, oder Grammatikkorpora definiert

sind. Den Hauptteil der Korpora bilden nach wie vor Zeitungskorpora. Die Korpora sind aber auch nach unterschiedlichen Projekten aufgeteilt wie das Mannheimer Korpus 1 und 2. Zusätzlich sind viele Biographien, Interviews und Protokolle, aber auch Fachtextkorpora vorhanden, wie es in der nächsten Abbildung (Abb.2) stichprobenartig veranschaulicht wird :

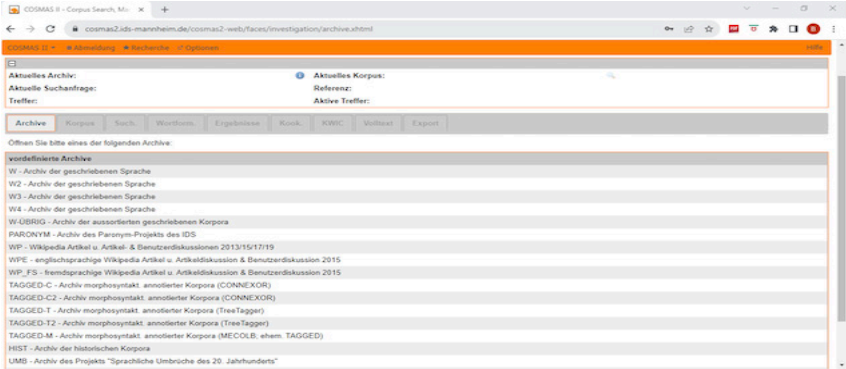


Abb. N° 2 : Teilkorpora des DeReKo

Für DeReKo wird keine ausgewogene oder gar « repräsentative » Zusammensetzung bezüglich Genres oder anderer Dimensionen angestrebt, wie auf den Abbildungen 2 und 3 ersichtlich ist ; denn, was als ausgewogen zu gelten hat, hängt letztlich von einer Forschungsfrage und einer zu untersuchenden Sprachdomäne ab. In der Abbildung 3 wurde das Fachsprachen-Korpus ausgewählt.

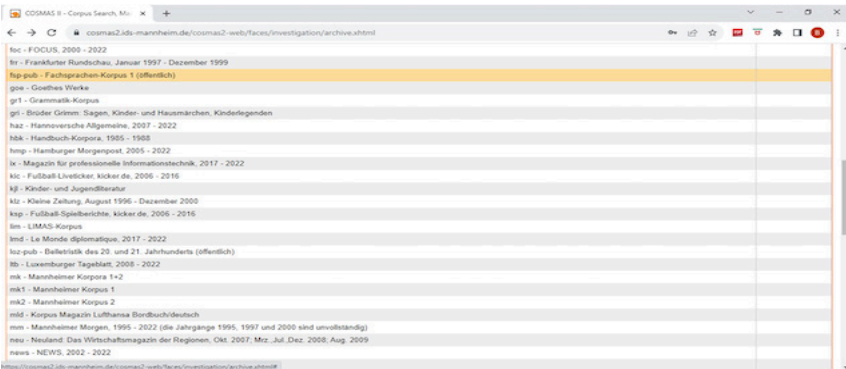


Abb. N° 3 : Fachsprachen-Korpus

Nachdem das Fachsprachen-Korpus ausgewählt wurde, kann man eine spezielle Suchanfrage starten. Die Auswahl des Korpus und die Formulierung der Suchanfrage sind für die Aussagekraft der Ergebnisse entscheidend. Es

bieten sich bei der Suchanfrage viele Möglichkeiten an. Es ist zu entscheiden, welcher sprachliche Begriff erfasst werden soll, welche Wortform gewählt wird, ob Groß-/Kleinschreibung relevant ist, ob ein gesamtes Paradigma gemeinsam betrachtet wird oder ob sogar Wortbildung berücksichtigt werden soll. Im Folgenden wurde beispielsweise das Wort « Wirtschaft » angefragt (Abb. 4)

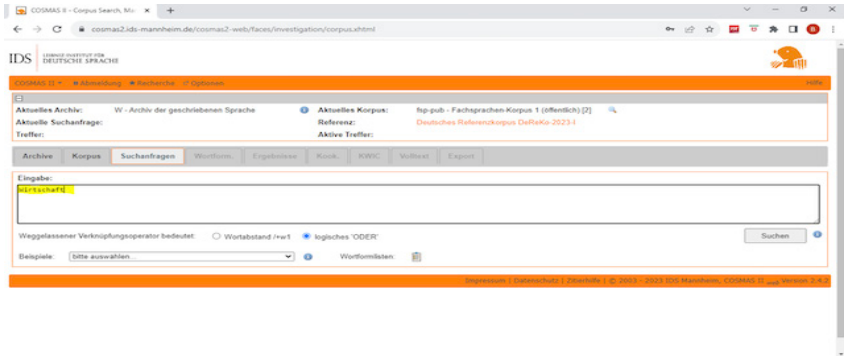


Abb. N° 4 : Suchanfrage « Wirtschaft »

Es wird dann eine KWIC-Ansicht (kurz für Key Words in Context ; auch : Konkordanz) aufgezeigt. Damit wird der Suchbegriff in der Mitte einer Textzeile dargestellt und grafisch hervorgehoben. Links und rechts davon wird so viel Kontext angegeben, wie die Zeile erlaubt (Vgl. Lemnitzer/Zinsmeier, 2006 : 196).

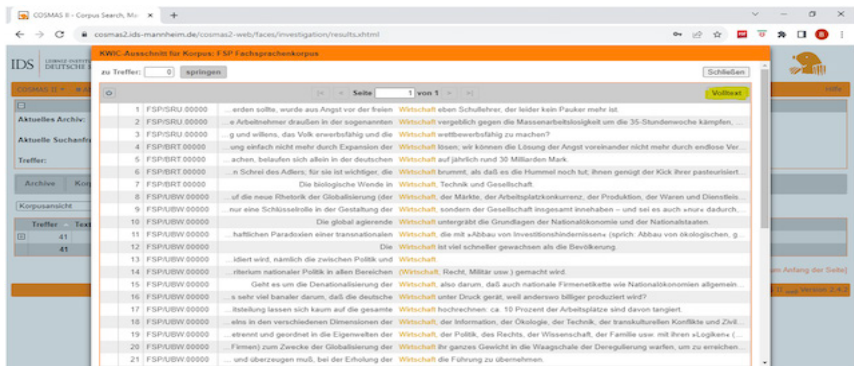


Abb. N° 5 : Eine KWIC-Ansicht

Als nächstes erscheint eine Vielzahl an Teiltextrn mit dem Inhalt « Wirtschaft ». Da viele Texte urheberrechtlich geschützt sind, ist im DeReKo der Zugriff auf vollständige Korpustexte, die in einer wissenschaftlichen Analyse gebraucht werden können, nicht immer möglich.

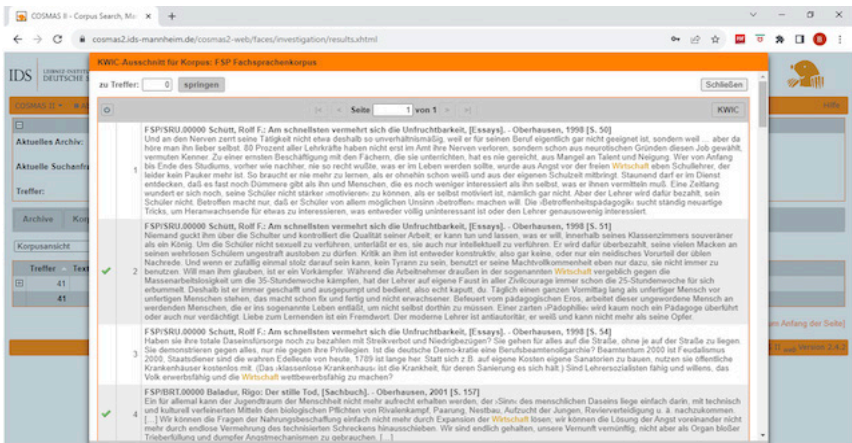


Abb.N° 6 : Textauswahl zum Wort « Wirtschaft »

Die Ergebnisse der Recherche können auch als virtuelles Korpus gespeichert werden, so dass der Zugriff darauf für weitere nachfolgende Untersuchungen gewährleistet ist. Es ist jedoch wichtig zu erwähnen, dass eine gewisse Methodenkompetenz und Einarbeitungszeit für die effiziente Nutzung von DeReKo notwendig sind (Vgl. Steyer, 2008: 189).

Schlussfolgerung

Korpuslinguistische Methoden gehören mittlerweile zum Standardrepertoire der sprachwissenschaftlichen Forschung. In der Korpuslinguistik wird unter anderem die deutsche Sprache in ihren verschiedenen Erscheinungsformen und ihrem Gebrauch untersucht. Das Leibniz-Institut für Deutsche Sprache (IDS) in Mannheim beinhaltet das größte Referenzkorpus weltweit. In diesem Beitrag wurde versucht, einen Einblick in die Korpuslinguistik, das Korpus und die Kriterien zur Erstellung eines Korpus zu geben. Weiterhin wurde das DeReKo beleuchtet und versucht, seine Ziele zu verdeutlichen. Das DeReKo dient primär dazu, die Verwendung der deutschen Sprache und ihre Entwicklung zu dokumentieren sowie eine praktische Grundlage für wissenschaftliche Analysen zur deutschen Gegenwartssprache zur Verfügung zu stellen, insbesondere für quantitative Untersuchungen, die sehr große Korpora erfordern. Des Weiteren wurde versucht aufzuzeigen, welche Schritte man verfolgen muss, um mit dem COSMAS II-System Einblick in das DeReKo zu gewinnen. Dadurch sollten Forscher und jedes Projekt selbst die Möglichkeit nutzen, aus DeReKo Korpora zu ziehen, die möglichst bezüglich ihrer jeweiligen Fragestellung repräsentativ sind. DeReKo ist eine sehr große Korpusressource für das

zeitgenössische Deutsch, die kontinuierlich erweitert und verbessert wird. Aktuelle Entwicklungen betreffen neue Lizenzmöglichkeiten, die Integration neuer Textarten und neue Möglichkeiten für Abfrage und Analyse von DeReKo.

Bibliografie

- Bopp, Sebastian. (2010) : *Einführung in die Korpuslinguistik mit DeReKo und COSMAS II*. [online] Abgerufen von URL : <https://colibris.link/caH0h> Zugriff am 11/10/2023.
- Bussmann, Hadumod (2008) : *Lexikon der Sprachwissenschaft*, 4., durchges. u. bibl. erg. Aufl. Körner Verlag, Stuttgart.
- Flinz, Carolina (2021) : Korpora als primäre Quellen von Tourlex. In : Taborek, Janusz ; Piosok, Michal & Woznicka, Marta (Hrsg.) : *Korpora in der Lexikographie*. *Lexicographica*. Series Maior. Berlin/Boston : de Gruyter.
- Flinz, Carolina & perkhun, Rainer (2018) : *Wortschatz und Kollokationen in ‚Allgemeine Reisebedingungen‘. Eine intralinguale und interlinguale Studie*. In : Čibej, Jaka ; Gorjanc, Vojko ; Kosem, Iztok & Krek, Simon (eds.) : *Proceedings of the XVIII EU-RALEX International Congress : Lexicography in Global Context*. Ljubljana University Press, Faculty of Arts : LjubljanaFlinz/ Perkuhn 2018 ; Flinz 2021
- Lemnitzer, Lothar / Zinsmeister, Heike (2006) : *Korpuslinguistik. Eine Einführung*. Tübingen : Narr.
- Lüngen, Harald (2017) : *DeReKo – Das Deutsche Referenzkorpus*, Schriftkorpora der deutschen Gegenwartssprache am Institut für Deutsche Sprache in Mannheim, De Gruyter.
- Mcenergy, Tony/ WILSON, Andrew (2003) : *Corpus Linguistics*. 2. Auflage, Edinburgh
- Scherer, Carmen (2006) : *Korpuslinguistik (Kurze Einführungen in die germanistische Linguistik - KEGLI)* Taschenbuch, 1. Edition, Universitätsverlag Winter GmbH Heidelberg.
- Steyer, Kathrin (2008) : *Kollokationen in deutschen Wörterbüchern und in der deutschen Wörterbuchforschung*. *Lexicographica : International Annual for Lexicography* 24.
- Wegera, Klaus-Peter/ Kwekkeboom, Sarah/ Herbers, Brigit/Schultz-Balluff, Schultz (2023) : *Sprachwandel : Kleinere Schriften zur deutschen Sprachgeschichte und Grammatikographie. Korpusbasierte Sprachdatengewinnung* (37-51), Erich Schmidt Verlag GmbH & Co. KG Berlin.

Internetquellen

- URL : <<https://www.ids-mannheim.de/>> Zugriff am 4/11/2023
- URL : <https://colibris.link/xnhRO> Zugriff am 11/10/2023
- URL : <https://colibris.link/Z0ZE8> Zugriff am 18/10/2023
- URL : <https://colibris.link/zhTN2> Zugriff am 17/11/2023
- URL : <https://colibris.link/8PCpT> Zugriff am 10/10/2023
- URL : <https://colibris.link/pTeC5> Zugriff am 2/10/2023
- URL : <https://colibris.link/T0jC3>

Zusammenfassung

In den letzten Jahren gab es eine hohe Anfrage nach verlässlichen Hilfsmitteln, die verhelfen eine wissenschaftliche Analyse zu erstellen. Dabei sind Korpora eine immer größere Grundlage insbesondere für quantitative Untersuchungen, die sehr große Korpora erfordern. Korpora sind aufbereitete Textsammlungen, die in unterschiedlichen Bereichen verwendet werden können und einen ganz konkreten Referenzbereich abdecken. Sie können zur Erforschung von sprachlichen Erscheinungen, Erstellung von Grammatiken, im Sprachenvergleich oder im Fremdsprachenunterricht aber auch zur Beantwortung bestimmter sprachwissenschaftlicher Untersuchungsfragen genutzt werden. In diesem Beitrag wird der Versuch unternommen, die praktische Bedeutung eines Korpus bzw. der Korpuslinguistik für sprachwissenschaftliche Analysen zu durchleuchten und darüber hinaus wird die Aufmerksamkeit auf die Korpuslandschaft DeReKo, die Ressourcen und Analysewerkzeuge sind, gelenkt und in den Mittelpunkt gerückt.

Stichwörter

Korpora, Korpuslandschaft DeReKo, Sprachwissenschaftliche Analyse, Leibniz-Institut für Deutsche Sprache (IDS)

مستخلص

في السنوات الأخيرة، كان هناك طلب كبير على الأدوات الموثوقة التي تساعد في إنشاء التحليل العلمي. تعد المجموعات الأساسية أساسًا متزايد الأهمية، خاصة بالنسبة للدراسات الكمية التي تتطلب مجموعات كبيرة جدًا. تعد المجموعات عبارة عن مجموعات نصية معدة يمكن استخدامها في مجالات مختلفة وتغطي منطقة مرجعية محددة للغاية. ويمكن استخدامها للبحث في الظواهر اللغوية، أو إنشاء القواعد، أو مقارنة اللغات أو تدريس اللغات الأجنبية، ولكن أيضًا للإجابة على بعض أسئلة البحث اللغوي. تحاول هذه المقالة تسليط الضوء على الأهمية العملية للمدونة أو لغويات المجموعة للتحليلات اللغوية، وبالإضافة إلى ذلك، تُلقت الانتباه إلى مشهد مجموعة DeReKo، وهي موارد وأدوات تحليل

كلمات مفتاحية

المجموعات المرجعية، المجموعة المرجعية لـ، التحليل اللغوي، معهد لايبنتز للغة الألمانية DeReKo، (IDS)

Résumé

Ces dernières années, il y a eu une forte demande pour des outils fiables qui aident à créer une analyse scientifique. Les corpus constituent une base de plus en plus importante, en particulier pour les études quantitatives qui nécessitent des corpus volumineux. Les corpus sont des collections de textes

préparés qui peuvent être utilisés dans différents domaines et qui couvrent un domaine de référence très spécifique. Ils peuvent être utilisés pour étudier les phénomènes linguistiques, créer des grammaires, comparer les langues ou enseigner les langues étrangères, mais aussi pour répondre à des questions de recherche linguistique spécifiques. Cet article tente de mettre en lumière l'importance pratique des corpus ou de la linguistique de corpus pour l'analyse linguistique et, en outre, attire l'attention sur le paysage de corpus DeReKo, qui sont des ressources et des outils d'analyse.

Mots-clés

Corpus, paysage de corpus DeReKo, analyse linguistique, Institut Leibniz pour la langue allemande (IDS)

Abstract

In recent years, there has been a high demand for reliable tools that help to create a scientific analysis. Corpora are an increasingly important basis, especially for quantitative studies that require extensive corpora. Corpora are prepared collections of text that can be used in different areas and cover a very specific reference area. They can be used to research linguistic phenomena, create grammar, compare languages, or teach foreign languages, but also to answer specific linguistic research questions. This article attempts to shed light on the practical significance of corpus or corpus linguistics for linguistic analysis and, in addition, draws attention to the DeReKo corpus landscape, which are resources and analysis tools.

Keywords

Corpora, corpus landscape DeReKo, linguistic analysis, Leibniz Institute for the German Language (IDS)