

III.2 Korpusanalyse

Viele, viele Daten, in denen Muster aufscheinen – die Korpusanalyse ist das Gegenstück zur Fallstudie. Dort ein einzelner Fall als Tiefenbohrung in einen Ausschnitt der Wirklichkeit, hier der breite, möglichst umfassende Überblick. Korpusanalysen in der Angewandten Linguistik zeigen, wie ganze Gemeinschaften und Gesellschaften Sprache nutzen. Dazu stellen die Forschenden einen Datenkörper zusammen – eben das Korpus –, den sie dann untersuchen nach der Häufigkeit bestimmter Muster des Sprachgebrauchs. Von dieser sprachlichen Oberfläche aus lässt sich schließen auf das Denken und Handeln von Gemeinschaften.

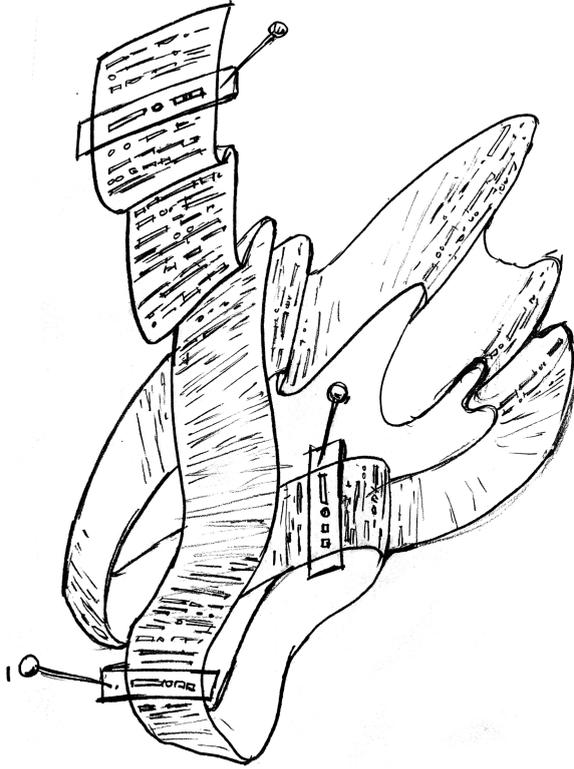
Das erste Kapitel [III.2.1](#) erklärt, wie Korpusanalysen Sprache und Kommunikation abbilden. Wenn wir Sprache und Medien nutzen, zum Beispiel im Alltag oder im Beruf, hinterlässt das Spuren: schriftliche Texte oder Ton- und Bilddateien. Daraus lassen sich Daten gewinnen zu geschriebener und gesprochener Sprache. Zusammengestellt nach bestimmten Regeln, bilden solche Daten ein Korpus authentischen, also wirklichen Sprachgebrauchs in einem bestimmten Anwendungsfeld. Ein solches Korpus eignet sich, um Muster wiederzufinden oder neu zu entdecken, die zeigen, wie Gemeinschaften ticken, wenn sie kommunizieren.

Das zweite Kapitel [III.2.2](#) steckt den Weg ab vom Datenmeer zum klaren Korpus. Jedes Korpus ist so angelegt, dass es die passenden Daten enthält, um bestimmte Forschungsfragen zu beantworten – Fragen etwa zum Energie- oder Klimadiskurs in einer bestimmten Region. Also sind diese Daten aufzuspüren, einzusammeln und systematisch abzulegen. So ist jedes Korpus ein Einzelstück, anders als alle andern – und doch haben sich im Lauf der Zeit Typen von Korpora herausgebildet, mit ähnlichem Zweck und Aufbau. Im Korpus selbst werden die Daten dann sozusagen etikettiert. Das hilft bei der Abfrage von Daten aus dem Korpus.

Im dritten Kapitel [III.2.3](#) lernen Sie die Methoden kennen, mit denen Sie Korpora abfragen. Das sind Verfahren, mit denen Sie Muster in den Daten des Korpus erkennen. Zuerst erfahren Sie, was man so alles zählen kann in einem Korpus, etwa das Vorkommen von Wörtern oder Wortgruppen. Wichtig ist aber auch, wie diese gezählten Objekte verteilt sind, etwa in einem Korpus, das einen Vergleich des Sprachgebrauchs über Zeiträume hinweg ermöglicht. Besonders interessant kann sein, welche Einheiten wie oft im Um-

feld von welchen anderen vorkommen – etwa das Wort *Flüchtling* in der Nähe von Begriffen wie *Welle* oder *Flut* oder *Integration*.

Im vierten Kapitel ^{III.2.4} erleben Sie, was Korpusanalysen nützen. Dies einerseits der Wissenschaft, die so mehr über den tatsächlichen Sprachgebrauch herausfindet. Andererseits profitiert aber auch die beforschte Praxis: Nicht immer sind sich die Menschen bewusst, wie hartnäckig sie als Individuen und Gemeinschaften bestimmte Muster des Sprachgebrauchs wiederholen, ob zielführende oder sinnwidrige. Wer mit Sprachkorpora arbeiten kann, gewinnt deshalb nicht nur als Forscher:in, sondern auch als reflektierte:r Praktiker:in in einem Sprach- und Kommunikationsberuf.



Julia Krasselt

III.2.1 Sprache und Kommunikation abbilden: Grundlagen der Korpuslinguistik

Korpora eignen sich, um wissenschaftlich fundierte Aussagen über Sprache und ihren tatsächlichen Gebrauch zu tätigen. Wie? – Das lernen Sie in diesem Kapitel: Es stellt die Korpuslinguistik als einen empirischen Zugang zu Sprache dar und zeigt Ihnen, in welchen Situationen in Alltag und Beruf Korpora eine entscheidende Rolle spielen. Für angehende Sprachprofis ist das Verständnis von Korpora als Zugang zu Sprachgebrauch wichtig, weil sie mit Korpusanalytischen Probleme der Sprach- und Kommunikationspraxis im Beruf lösen können.

Haltung: Sie entwickeln ein Verständnis dafür, dass linguistisch annotierte Korpora einen zentralen Zugang zu Sprache und ihrem Gebrauch darstellen und sich deswegen für Forschungs- und Praxisaufgaben eignen. **Wissen:** Sie schärfen wichtige methodische Konzepte für Ihre Arbeit mit Korpora. **Können:** Dieses Methodenwissen können Sie in ausgewählten Aufträgen für Korpusanalysen anwenden.

Lernziele

Das Kapitel vertieft die Fachbegriffe ^{a, b}, die zu Beginn dieses Bandes eingeführt wurden ^{1.1,2}, und zeigt, wo Korpora in Alltag, Beruf und Forschung eingesetzt werden ^c. Abschließend behandelt es zwei Vorgehensweisen, die kennzeichnend für das Forschen mit Korpora sind ^d.

Aufbau

- a Korpus und Korpuslinguistik:
Sprachgebrauch in authentischen Kontexten
- b Primärdaten, Annotationen, Metadaten:
Bestandteile linguistischer Korpora
- c Wörterbücher, Sprachkompetenz, maschinelle Übersetzung:
Korpora in Alltag und Beruf
- d Korpusgetrieben oder korpusbasiert:
Forschen mit und über Korpora

Prof. Dr. Julia Krasselt hat langjährige Forschungserfahrung in den Bereichen Digitale Linguistik, Korpuslinguistik und Diskursanalyse.

Autorin

Sie leitet das digitale Textkorpus *Swiss-AL*. In der Lehre verbindet sie die Korpuslinguistik mit diskurslinguistischen Fragestellungen.

<https://www.zhaw.ch/de/ueber-uns/person/krss/>

Thema 2.1.a Korpus und Korpuslinguistik: Sprachgebrauch in authentischen Kontexten

Stellen Sie sich das folgende Szenario vor: Sie arbeiten bei einem großen Verein, der sich sozialpolitisch engagiert. Sie sind Teil einer Arbeitsgruppe, die eine neue landesweite Informationskampagne zum Thema *Sozialhilfe* erarbeitet. Neben Ihnen sind weitere Übersetzer:innen, Integrationsexpert:innen und Expert:innen aus Kommunikation sowie Marketing Teil der Arbeitsgruppe. Sie sollen herausfinden, welchen Stellenwert das Thema *Sozialhilfe* in den Massenmedien hat, wer sich noch dazu öffentlich äußert und was die häufigsten Bezeichnungen für die betroffenen Personengruppen in allen Landessprachen sind. – Wie können Sie diese komplexe Aufgabe angehen?

Einen geeigneten Zugang stellt die Analyse öffentlicher Äußerungen dar. Das heißt, Sie benötigen Zugang zum empirisch beobachtbaren Sprachgebrauch in authentischen Kontexten zum Thema *Sozialhilfe*. Einen solchen Zugang zum Sprachgebrauch bietet die Korpuslinguistik ^{1.1.2.a}. Sie arbeitet mit umfangreichen, digital vorliegenden Sammlungen von sprachlichen Äußerungen. Was in einer solchen Sammlung – einem Korpus – enthalten ist, hängt vom Erkenntnisinteresse sowie von Fragen aus Forschung und Praxis ab.

So kann ein Korpus beispielsweise aus Zeitungsartikeln zum Thema *Sozialhilfe* bestehen. Ein solches Korpus kann aufgrund seiner Größe nicht mehr manuell analysiert werden, etwa durch Lesen und Analysieren aller enthaltenen Artikel. Stattdessen werden in der Korpuslinguistik spezifische Tools und Methoden zur Analyse von Korpora eingesetzt ^{1.1.2.d}. Diese ermöglichen dann beispielsweise das Erfassen der Häufigkeit von Wörtern wie *Sozialhilfebezügler* oder *Sozialhilfeempfängerin*. Zudem helfen die spezifischen Tools und Methoden dabei, sich die Kontexte, in denen diese Wörter verwendet werden, systematisch anzuschauen.

Die Korpuslinguistik stellt also in erster Linie einen quantitativen Zugang zu Sprachgebrauch dar: Es geht vor allem darum, Typi-

sches und Frequentes, also Häufiges, im Sprachgebrauch zu identifizieren. Weniger zentral hingegen sind die Beobachtung und Beschreibung von Einzelfällen. Die Korpuslinguistik ist interessiert an Sprachgebrauchsmustern. Sie fragt danach, welche sprachlichen Einheiten immer wieder – also musterhaft – verwendet werden. Solche Einheiten können zum Beispiel Wörter, Wortgruppen und syntaktische Konstruktionen sein, aber auch Metaphern, rhetorische Figuren und Argumentationen.

In der Korpuslinguistik spielen deswegen die Angabe von Häufigkeiten sowie die Verwendung statistischer Maßzahlen eine wichtige Rolle. Typisch sind beispielsweise Formulierungen wie die folgende: *Das Wort Sozialhilfeempfänger tritt signifikant häufig im Kontext von Sanktion und Kontrolle auf.* Korpuslinguistische Zugänge zu Sprache sind daran interessiert, solche Befunde zu erklären und mit sprach-, kognitions- und kommunikationswissenschaftlichen Theorien zu verknüpfen. Das heißt, dass Korpuslinguistik immer mehr ist als die reine Quantifizierung von Sprachgebrauchsmustern. Zum Beispiel lässt sich die Beobachtung zu typischen Kontexten des Wortes *Sozialhilfeempfänger* so interpretieren, dass die Gruppe der Personen, die Sozialhilfe empfängt, einseitig auf einen Aspekt reduziert wird: den der Kontrolle und Sanktion.

Skizzieren Sie den Zusammenhang zwischen Korpuslinguistik und Sprachgebrauch.

Schnelltest

Auf der Webseite zum Buch finden Sie etwa die folgende weiterführende Übung: Sie suchen in einem Korpus das Wort *Sozialhilfeempfänger* und erkunden sein sprachliches Umfeld.

Training

Thema 2.1.b Primärdaten, Annotationen, Metadaten: Bestandteile linguistischer Korpora

Wenn ein Korpus eine digital vorliegende Sammlung von Sprachdaten ist ^{III.2.1.a} – wäre dann bereits ein digitales Archiv mit allen Medienmitteilungen einer Bundesbehörde ein Korpus? Oder können die riesigen Textmengen, die das US-amerikanische Unternehmen OpenAI zum Trainieren seines Chat-Bots ChatGPT verwendet, als Korpus bezeichnet werden? Die kurze Antwort lautet: nein. Denn: Ein Korpus ist eine Sammlung von Sprachdaten, die nach spezifischen und transparenten Kriterien zusammengestellt ist und

für den Zweck der empirisch basierten linguistischen, kommunika-tions- und medienwissenschaftlichen Forschung aufbereitet wird.

Was heißt das genau? Bereits im ersten Teil des Buches haben Sie gelernt, dass man erst dann von einem Korpus spricht, wenn es neben Primärdaten über Annotationen und Metadaten verfügt ^{I.1.2.c.} Schauen wir uns das etwas vertiefter an:

Als Primärdaten werden natürlichsprachliche Daten ganz unterschiedlichen Ursprungs bezeichnet; sie stellen das Untersuchungsobjekt korpuslinguistischer Studien dar ^{I.1.2.c.} Solche Primärdaten können – je nach Untersuchungsinteresse in Forschung und Praxis – zum Beispiel Medienmitteilungen eines Unternehmens sein, Artikel aus Tages- und Wochenzeitungen, Gespräche zwischen Ärzt:innen und Patient:innen oder auch Youtube-Videos samt Kommentaren. Primärdaten können also schriftlicher, mündlicher oder multimodaler Natur sein und entweder bereits digital oder noch analog vorliegen. Oftmals müssen die Primärdaten eines Korpus noch bearbeitet werden, bevor sie schließlich Teil des Korpus werden; man nennt diesen Schritt *Preprocessing*. Ein Beispiel wäre die Entfernung typografischer Merkmale in Medienmitteilungen oder die Transkription von Gesprächsdaten.

Metadaten haben Sie als *Daten über Daten* kennengelernt ^{I.1.2.c.}; sie beschreiben die Primärdaten des Korpus näher. Mit ihnen kann beispielsweise erfasst werden, in welcher Sprache ein Text verfasst wurde, von welcher Autorin oder zu welchem Zeitpunkt. Damit ermöglichen Metadaten Aussagen über Teilmengen eines Korpus, beispielsweise über die Frequenz eines Wortes in zwei verschiedenen Jahren ^{III.2.3.b.}

Unter Annotationen versteht man die Anreicherung des Korpus mit linguistischen Merkmalen der Einheiten des Korpus ^{I.1.2.c.} Solche Merkmale betreffen alle Sprachebenen, zum Beispiel die Morphologie, Syntax, Semantik oder Pragmatik. Annotationen leisten, dass diese Merkmale explizit und somit systematisch auffindbar und auszählbar gemacht werden. Nur durch Annotationen wird es zum Beispiel möglich, systematisch alle in einem Korpus vorkommenden Substantive oder Organisationsnamen auszuzählen ^{III.2.3.a.}

Schnelltest Erklären Sie die Funktionen von Metadaten und Annotationen in Korpora.

Training Auf der Webseite zum Buch finden Sie zum Beispiel einen kleinen Test, mit dem Sie üben können, zwischen Annotationen und Metadaten zu unterscheiden.

Thema 2.1.c Wörterbücher, Sprachkompetenz, maschinelle Übersetzung: Korpora in Alltag und Beruf

Der Begriff *Korpus* ist im Kontext von Alltag und Beruf kaum bekannt. Dennoch gibt es zahlreiche Situationen, in denen wir uns Praktiken und Tools bedienen, die auf sehr großen Sammlungen von Sprachdaten beruhen.

Wörterbücher wie der Duden basieren auf sehr großen Korpora. Lexikograf:innen verwenden diese Korpora, um Lexikoneinträge für Wörter zu erstellen. Angaben zur Bedeutung und Häufigkeit eines Wortes basieren also auf der Analyse zahlreicher Belegstellen im Korpus. Historisch gesehen kommt die Korpuslinguistik genau aus diesem Bereich: Die ersten Korpora wurden mit dem Ziel erstellt, Wörterbücher zu schreiben. Dahinter liegt die Erkenntnis, dass nur durch Beobachtung eines Wortes in seinem Gebrauch etwas über seine Bedeutung herausgefunden werden kann III.2.3.c/d.

Ein weiteres Beispiel zur Nutzung von Korpora in Alltag und Beruf liegt im Bereich Bildung – insbesondere beim Sprachenlernen und beim Erwerb von Schreibkompetenzen. Viele Lehrmaterialien werden heute korpusbasiert erstellt. Dazu gehört die Angabe zu Frequenzen und typischen Gebrauchskontexten beim Erlernen von neuem Vokabular. Dies gilt auch für Schreibassistenzsysteme, die beim Schreiben von Texten Formulierungsvorschläge machen sowie Grammatik und Rechtschreibung überprüfen – und sich dabei meist an die Variante halten, die im Korpus am häufigsten vorkommt.

Neben solchen tatsächlich auf linguistischen Korpora beruhenden Praktiken gibt es weitere Praktiken, die ebenfalls auf der Sammlung von Textdaten beruhen. In der Organisationskommunikation üblich ist beispielsweise die Medienbeobachtung: Journalistische Beiträge werden systematisch anhand von Suchbegriffen gesammelt und regelmäßig ausgewertet.

Eine weitere Praktik, die mittlerweile nahezu alltäglich geworden ist, ist die Verwendung von Systemen für die maschinelle Übersetzung wie DeepL oder Google Translate. Dass diese Systeme inzwischen sehr gute Übersetzungen erzeugen, liegt daran, dass sie kontinuierlich aus der Beobachtung großer Textmengen lernen. Auf gleiche Weise funktionieren auch Sprachassistenzsysteme wie Alexa und Siri oder Textgenerierungssysteme wie ChatGPT oder Gemini.

Solche Entwicklungen im Bereich der Sprachtechnologie und des maschinellen Lernens sind prominente Beispiele für künstliche Intelligenz, die auf riesigen Datenmengen trainiert wird. Die verwendeten Daten unterscheiden sich aber in wichtigen Punkten von Korpora III.2.1.b: Hinter Systemen wie DeepL, Alexa und ChatGPT stehen zwar auch große Mengen von Sprachdaten; deren Zusammenstellung ist aber nicht transparent und inhaltlich begründet. Auch sind diese Daten nicht abfragbar. Annotationen und Metadaten spielen, wenn überhaupt, nur eine marginale Rolle.

Linguistische Korpora dagegen folgen sehr ausgewählten, inhaltlich begründbaren und vor allem transparenten Modellierungskriterien III.2.2.a, die durch die Forschenden in Abhängigkeit vom jeweiligen Erkenntnisinteresse definiert werden. Korpora und die mit ihnen durchgeführten Analysen erfüllen somit zentrale wissenschaftliche Gütekriterien. Also lassen sich aus Korpusanalysen zuverlässige und gültige Aussagen für die Sprachpraxis ableiten.

- Schnelltest Erklären Sie, worin sich KI-Textsammlungen von linguistischen Korpora unterscheiden.
- Training Auf der Webseite zum Buch finden Sie eine Zusammenstellung von Korpora und Tools, die Sie in Alltag und Beruf einsetzen können.

Thema 2.1.d Korpusgetrieben oder korpusbasiert: Forschen mit und über Korpora

Korpora spielen eine wichtige Rolle in der angewandten linguistischen Forschung. Korpusbasierte Wörterbücher, Schreibassistenzsysteme und Lern- und Lehrmittel sind Beispiele für Errungenschaften der Forschung mit und über Korpora. Was aber kennzeichnet die Rolle, die Korpora in der Angewandten Linguistik spielen?

In der Angewandten Linguistik wird dem Forschen mit und über Korpora ein unterschiedlicher Status beigemessen. Anders gesagt: Es gibt unterschiedliche Auffassungen darüber, ob die Korpuslinguistik nur einen Werkzeugkasten zur Analyse von Sprachgebrauch bereitstellt oder aber eine eigenständige Disziplin ist. Was heißt das genau?

Betrachtet man die Korpuslinguistik als Werkzeugkasten, stehen konkrete Hypothesen im Vordergrund, die mithilfe von Korpora

überprüft werden, aber schon vor dieser Überprüfung bestehen. Beispielsweise kann man die Hypothese überprüfen, dass in Deutschschweizer Medien das generische Maskulinum häufiger verwendet wird als neutrale oder gegenderte Formen.

Wichtig: Diese Hypothese besteht bereits *vor* der Konsultation eines Korpus, beispielsweise aufgrund der eigenen Medienlektüre und somit subjektiver Eindrücke. Man schlägt dann in einem Korpus Deutschschweizer Medien gezielt Bezeichnungen für Personen und Personengruppen nach, beispielsweise *Ärzte, Ärztinnen, Ärztinnen und Ärzte, Ärzt:innen* etc. Aufgrund von Häufigkeitsverteilungen kann man dann die anfängliche Hypothese entweder bestätigen oder aber auch ablehnen (*falsifizieren*).

Diese Art der Korpusnutzung haben Sie bereits unter der Bezeichnung *corpus-based* kennengelernt ^{I.1.2.b}. Korpora sind dann in erster Linie reine Nachschlagewerke. Einen anderen Status haben Korpora in Untersuchungen, die als *corpus-driven* bezeichnet werden ^{I.1.2.b}. Hier werden vorab keine Hypothesen über Sprachgebrauch formuliert. Stattdessen werden Hypothesen erst aufgrund von Korpusanalysen formuliert und weiterverfolgt.

Corpus-driven wäre es, in einem Korpus zunächst einmal alle Bezeichnungen und Schreibweisen für Berufe zu identifizieren. Man stellt dann beispielsweise fest, dass *Politiker* häufiger vorkommt als *Politikerinnen*. Dann kann man sich die Kontexte dieser Berufsbezeichnungen anschauen und stellt fest, dass die Bezeichnung *Politikerinnen* häufiger im Kontext von Wörtern wie *jung* oder *Mutter* vorkommt als die Bezeichnung *Politiker* ^{III.2.3.d}. Man kann so zu Hypothesen über Stereotypisierungen von Berufen kommen. Heißt: Es wurde weder explizit nach *Politiker* und *Politikerinnen* gesucht noch nach Kontexten, in denen *Politikerinnen* mit dem Wort *jung* oder *Mutter* auftritt. Stattdessen hat man sich vom Korpus und geeigneten Methoden (hier: Frequenz- und Kontextanalyse) leiten lassen.

Die Stärke eines solchen Vorgehens sollte deutlich geworden sein: Die Arbeit mit Korpora eröffnet bei einem Vorgehen, das corpus-driven ist, eine vollkommen neue Perspektive auf Sprache: Unbekannte Muster und Strukturen können entdeckt werden.

Arbeiten Sie korpusgestützt oder -basiert, wenn Sie prüfen, ob *Natel* oder *Handy* häufiger ist?

Schnelltest

Auf der Webseite zum Buch finden Sie weiterführende Aufgaben, etwa zusätzliche Beispiele, die Sie dem korpusbasierten und korpusgestützten Vorgehen zuordnen sollen.

Training

Julia Krasselt

III.2.2 Die Qual der Wahl: Korpora auswählen und zusammenstellen

Bevor Sie mit einem Korpus arbeiten können, müssen Sie sich zunächst damit auseinandersetzen, welche Daten es enthalten soll, damit Sie eine wissenschaftliche Fragestellung oder ein Praxisproblem sinnvoll bearbeiten können. Korpora stehen immer in einem spezifischen Verhältnis zum Gegenstand, der mit ihnen untersucht werden soll – etwa dem aktuellen Diskurs zur Klimaerwärmung. Wie dieses Verhältnis beschaffen ist und was es bei der Wahl eines geeigneten Korpus zu beachten gibt, erfahren Sie in diesem Kapitel.

Lernziele **Haltung:** Sie verstehen, dass Korpora immer *Modelle* eines linguistischen Untersuchungsgegenstandes darstellen, etwa eines thematischen Diskurses. **Wissen:** Sie lernen Typen von Korpora und erste Techniken der Korpusnutzung kennen. **Können:** Dieses Wissen wenden Sie in konkreten Übungen an, in denen Sie mit einem vorhandenen Korpus arbeiten.

Aufbau Das Kapitel widmet sich der Zusammenstellung, Auswahl und Erstellung von Korpora. Es erklärt, was es bei der Wahl eines geeigneten Korpus zu beachten gilt ^a und welche Korpus Typen es gibt ^b. Schließlich steht das konkrete Arbeiten mit verfügbaren Korpora im Vordergrund: die Herstellung und Nutzung von Annotationen ^c sowie die Nutzung von Korpora in der Analysesoftware CQPweb ^d.

- a Repräsentativität, Modellierung, Sampling:
Welche Daten soll ein Korpus enthalten?
- b Mehrsprachig, synchron, spezialisiert:
Typen linguistischer Korpora
- c Annotationen und Tagsets:
Hinter die Sprachoberfläche schauen
- d Subkorpora und Suchanfragen:
Korpora und Tools richtig nutzen

Autorin Prof. Dr. Julia Krasselt III.2.1

Thema 2.2.a Repräsentativität, Modellierung, Sampling: Welche Daten soll ein Korpus enthalten?

Am Beginn jeder korpuslinguistischen Untersuchung steht die Zusammenstellung und Aufbereitung geeigneter Daten. Interessiert beispielsweise der öffentliche Schweizer Diskurs über gendergerechte Sprache, dann sind Artikel journalistischer Medien, Beiträge auf Social-Media-Plattformen oder Angebote der Organisationskommunikation relevant. Nicht von Interesse ist hingegen, wie man sich im Privaten über dieses Thema unterhält – es geht ja eben um den öffentlichen Diskurs.

Ein Korpus soll valide, also gültige Aussagen über einen spezifischen Sprachausschnitt ermöglichen. Man sagt, dass das Korpus für diesen Sprachausschnitt repräsentativ sein soll. Damit ist gemeint, dass die Befunde und Schlussfolgerungen der Analyse generalisierbar sind und nicht nur für das Korpus gelten, sondern generell für den untersuchten Sprachausschnitt, etwa für das öffentliche Sprechen über gendergerechte Sprache ^{1.1.2.c}.

Warum ist diese Überlegung wichtig? Korpora enthalten in seltenen Fällen alle produzierten Äußerungen, sondern immer nur eine Auswahl. Obwohl man allein aus praktischen und rechtlichen Gründen niemals alle je geschriebenen Medienartikel, Social-Media-Posts und Webtexte von Organisationen zu gendergerechter Sprache in ein Korpus aufnehmen kann, sollen dennoch Aussagen über den öffentlichen Diskurs zum Thema getroffen werden. Es müssen also vor der Zusammenstellung des Korpus Parameter definiert werden, die eine Auswahl der zu untersuchenden Daten begründen. Diesen Vorgang bezeichnet man als Sampling oder auch als Modellierung: Das Korpus ist das Modell für die Untersuchung einer spezifischen Fragestellung zu einem Forschungsgegenstand ^{Dreesen & Stücheli-Herlach, 2019}.

Dafür muss zunächst überlegt werden, welche Eigenschaften des Untersuchungsgegenstandes das Modell enthalten muss, um die praxisbezogene Fragestellung untersuchen zu können. Man kann sich das wie das Bauen eines Legohauses vorstellen: Es wird zwar niemals genau so gebaut sein wie ein echtes Haus; es würden zum Beispiel Steckdosen und Stromkabel fehlen. Dennoch könnte man es anhand typischer Merkmale wie Wände oder Fenster als Haus identifizieren. Zudem wird es so beschaffen sein, wie es dem Zweck der Spieler:in dienlich ist, etwa mit besonders großen Bäumen darum herum.

Wichtige Parameter in Bezug auf öffentliche Diskurse über gendergerechte Sprache sind sicherlich Sprache, Urheber:in des Texts und Suchworte. Zum Beispiel sollte unser Modell in geeigneter Form Mehrsprachigkeit sowie die Medien- und Unternehmenslandschaft abbilden. Außerdem sollte das Korpus Texte enthalten, in denen es tatsächlich um gendergerechte Sprache geht, zum Beispiel weil die Wörter *gendergerecht* oder *Gender-Stern* darin vorkommen.

Es gibt Korpora, die mit dem Anspruch erstellt werden, repräsentativ für eine ganze Sprache zu sein – und nicht nur für einen spezifischen Sprachausschnitt, wie bisher erläutert. Solche Korpora werden als Referenzkorpora bezeichnet. Ein Beispiel ist das British National Corpus (BNC), ein Referenzkorpus für britisches Englisch im späten 20. Jahrhundert BNC Consortium, 2007. Es enthält sowohl gesprochene als auch geschriebene Sprache und ist hinsichtlich Textsorten und Textlängen ausgewogen 1.1.2.c: Die enthaltenen Texte sind von vergleichbarer Länge und auch einzelne Textsorten wie fiktionale Texte und Zeitungsartikel sind in vergleichbarer Zahl repräsentiert.

- Schnelltest Erklären Sie, was *Modellierung* in der Korpuslinguistik bedeutet und warum sie wichtig ist.
- Training Auf der Webseite zum Buch finden Sie unter anderem zwei Szenarien, für die Sie sich ein geeignetes Korpusmodell überlegen sollen.

Thema 2.2.b Mehrsprachig, synchron, spezialisiert: Typen linguistischer Korpora

Die Korpuslinguistik kennt mehrere Typen von Korpora. Für welchen Typ man sich entscheidet, hängt grundsätzlich vom Modell ab, das man für den Untersuchungszweck benötigt III.2.2.a. Eine grundlegende Unterscheidung bezieht sich auf das Kommunikationsmedium: Schriftsprachliche Korpora enthalten geschriebene Texte, mündliche Korpora enthalten gesprochene Äußerungen (die transkribiert worden sind) und multimodale Korpora enthalten weitere Elemente wie Ton und Bild.

Eine weitere Unterscheidung betrifft Korpora, die Sprachdaten in nur einer Sprache enthalten, und Korpora, die Sprachdaten in mehreren Sprachen enthalten. Bei mehrsprachigen Korpora können die Sprachdaten in den unterschiedlichen Sprachen entweder

vergleichbar oder parallel sein. Sie kommen zum Beispiel in der Translations- und Dolmetschwissenschaft zum Einsatz. Ein Parallelkorpus besteht aus Originaltexten sowie Übersetzungen in mindestens einer weiteren Sprache. Ein besonderes Merkmal von Parallelkorpora ist die explizite Zuordnung voneinander entsprechenden Segmente, etwa Sätzen. So können zum Beispiel Übersetzungsvarianten eines Wortes untersucht werden. Bei einem Vergleichskorpus hingegen stehen die enthaltenen Daten in einem Ähnlichkeitsverhältnis, weil sie eventuell das gleiche Thema betreffen oder die gleiche Textsorte repräsentieren. Sie sind aber keine Übersetzungen voneinander.

Korpora unterscheiden sich außerdem hinsichtlich ihrer zeitlichen Erstreckung. Synchrone Korpora enthalten Sprachdaten aus einem bestimmten, kürzeren, in sich abgeschlossenen Zeitraum. Diachrone Korpora decken einen längeren, in mehrere Phasen unterteilbaren Zeitraum ab. Sie ermöglichen deswegen die Untersuchung von Wandelphänomenen, etwa auf der Ebene gesellschaftlicher Diskurse. Will man den Migrationsdiskurs eines Landes untersuchen, kann man so bei der Modellierung unterschiedliche Migrationsphasen berücksichtigen.

Schließlich wird zwischen allgemeinen und Spezialkorpora unterschieden. Einen Typus des allgemeinen Korpus haben Sie bereits in Form des Referenzkorpus kennengelernt III.2.2.a. Eine gängige Auffassung ist, dass Referenzkorpora eine Sprache möglichst umfassend erfassen, indem sie, wie im Falle des BNC, verschiedene Textsorten und Erscheinungsformen einbeziehen, zum Beispiel Belletristik und Blogs III.2.2.a. Sie sollen verallgemeinerbare Aussagen ermöglichen, etwa über das britische Englisch des 20. Jahrhunderts. Spezialkorpora hingegen repräsentieren spezifischere Ausschnitte des Sprachgebrauchs. Ein wichtiger Vertreter sind sogenannte Lernendenkorpora. Sie enthalten Sprachdaten von Sprachlernenden und verfügen so unter anderem über Annotationen III.2.2.c lernensprachlicher Merkmale. Aber auch Korpora zu thematischen Diskursen wie dem Energiediskurs gehören zu den Spezialkorpora.

Korpora lassen sich mithilfe dieser Dimensionen näher bestimmen. Die genannten Dimensionen sind aber in der Praxis nicht immer trennscharf. Oft stehen bei einem Korpus eine oder zwei Dimensionen im Vordergrund. Beispielsweise ist die an der ZHAW Zürcher Hochschule für Angewandte Wissenschaften entwickelte Korpusfamilie Swiss-AL nach synchronen und schriftsprachlichen

Parametern modelliert ^{Krasselt et al., 2023}. Andere Beispiele, die die Anwendung der genannten Dimensionen zeigen, sind mündliche Lernendenkorpora, diachrone Parallelkorpora oder multimodale Spezialkorpora.

- Schnelltest Benennen Sie fünf Korpusstypen und erklären Sie die Unterschiede.
- Training Auf der Webseite zum Buch finden Sie als weiterführende Aufgaben zum Beispiel Beschreibungen von vier existierenden Korpora, die Sie den Korpusstypen zuordnen.

Thema 2.2.c Annotationen und Tagsets: Hinter die Sprachoberfläche schauen

Korpora erhalten ihren Wert insbesondere dadurch, dass sie nicht nur Primärdaten enthalten, sondern weitere linguistische Informationen ^{III.2.1.b}. Solche Informationen werden als Annotationen bezeichnet ^{I.1.2.c}. Sie beziehen sich auf die linguistischen Merkmale der Primärdaten des Korpus. Der Vorgang des Auszeichnens mit linguistischer Information wird in der Korpuslinguistik als Annotieren oder auch als Taggen bezeichnet.

Merkmale, die annotiert werden können, betreffen ganz unterschiedliche Ebenen des Sprachsystems. So kann jedes Wort eines Korpus mit einer Wortart sowie einer Grundform – dem sogenannten Lemma – annotiert sein ^{III.2.3.a}. Interessiert man sich zum Beispiel dafür, welche Adjektive häufig das Wort *Integration* attribuieren, lässt sich das mithilfe von Wortart-Annotationen sehr leicht herausfinden: Im Korpus kann gezielt nach Adjektiven gesucht werden, auf die das Wort *Integration* folgt, etwa *gelungene Integration*, *schwierige Integration*. Auf der Ebene der Syntax hingegen kann ein Wort mit seiner syntaktischen Funktion, die es im Satz übernimmt, annotiert sein. Das ermöglicht zum Beispiel, alle Belege im Korpus zu finden, in denen *Integration* die Funktion des Subjekts im Satz übernimmt, etwa *Integration fördert den Zusammenhalt zwischen den Menschen*.

Annotationen können entweder mithilfe geeigneter Tools – sogenannter Tagger – automatisiert erzeugt oder aber auch manuell erstellt werden. Die Annotation von Wortart und Lemma geschieht meist automatisiert, die meisten Korpora verfügen standardmäßig über diese beiden Annotationsebenen. Für die Annotation werden

üblicherweise standardisierte Tagsets verwendet. Damit sind vordefinierte Listen von Wortarten gemeint, für die spezifische Abkürzungen – sogenannte Tags – verwendet werden. Für das Deutsche üblich ist das Stuttgart-Tübingen-Tagset, kurz: STTS Schiller et al., 1999. So erhalten zum Beispiel alle attributiven Adjektive den Tag *ADJA* und alle appellativen Nomen den Tag *NN*.

Annotationen leisten zwei entscheidende Dinge: Zum einen ermöglichen sie, dass man in einem Korpus explizit nach linguistischen Kategorien suchen kann, die an der Wortoberfläche schlicht nicht erkennbar sind – oft sieht man einem Wort nicht an, ob es Adjektiv oder Adverb ist. Zum anderen ermöglichen Annotationen eine Disambiguierung: Wörter, die an der Oberfläche eine identische Form haben, können über linguistisch vollkommen verschiedene Merkmale verfügen. So ist das Wort *einen* in den folgenden beiden Sätzen mal ein Verb und mal ein unbestimmter Artikel:

- Sie wollen das Land wieder einen.
- Er kauft sich einen Stift.

Annotationen müssen manuell erstellt werden, wenn es keine gut funktionierenden automatisierten Tools gibt. Eine manuelle Annotation erfolgt typischerweise durch mehrere Personen. Dann kann ein sogenanntes Inter-Annotator-Agreement angegeben werden: Es drückt aus, zu welchem Grad mehrere Personen bei der Vergabe von Annotationen übereinstimmen. Bei der manuellen Annotation stehen meist keine standardisierten Tagsets bereit, sondern es werden sehr projektspezifische Tags und Guidelines zur Vergabe dieser Tags entwickelt. Ein Beispiel sind Fehlerannotationen in Lernendenkorpora.

Erklären Sie, worin der Nutzen von Annotationen besteht.

Schnelltest

Auf der Webseite zum Buch finden Sie an weiterführenden Aufgaben etwa einen Screencast und eine Übung zum Abfragen von Annotationsebenen in der Swiss-AL-Korpusfamilie.

Training

Thema 2.2.d Subkorpora und Suchanfragen: Korpora und Tools richtig nutzen

Vor Ihnen liegt die folgende Fragestellung: *Beeinflusst die politische Orientierung eines Akteurs, wie er öffentlich über die Themen Klima-*

wandel und Nachhaltigkeit spricht? Noch konkreter: Welche Muster des Sprachgebrauchs sind typisch für einzelne Parteien, wenn sie sich beispielsweise im Parlament zu diesen Themen zu Wort melden?

Oftmals eignen sich existierende Korpora, um eine Fragestellung wie diese zu bearbeiten, weil sie dem benötigten Modell bereits sehr nahekommen ^{III.2.2.a}. In anderen Fällen wiederum ist es notwendig, selbst ein Korpus zu erstellen, in dem die benötigten Primärdaten gesammelt und aufbereitet werden. Am Beispiel der Korpusfamilie Swiss-AL lernen Sie im Folgenden, wie Teile eines bereits existierenden Korpus genutzt werden können.

Ein geeignetes Korpusmodell für das oben skizzierte Fallbeispiel enthält Redebeiträge von Politiker:innen im Parlament. Für jeden Redebeitrag muss die Parteizugehörigkeit der redenden Person bekannt sein. Mit dem Parlamentskorpus der Swiss-AL-Korpusfamilie steht Ihnen ein Korpus zur Verfügung, das diesen Modellierungskriterien entspricht: Es enthält das Amtliche Bulletin, d. h. die Wortprotokolle des Schweizer National- und Ständerats.

Dieses Parlamentskorpus ist über das Korpusanalysetool *CQPweb* ^{Hardie, 2012} zugänglich, eines der am weitest verbreiteten Tools für Korpusanalysen ^{III.2.3}. Insbesondere ermöglicht CQPweb, nur in spezifischen Teilen eines eigentlich viel größeren Korpus zu suchen ^{I.1.2.d}. In anderen Worten: CQPweb erlaubt es, für eine bestimmte Untersuchung sogenannte Sub- oder Teilkorpora zu erstellen. Für das skizzierte Fallbeispiel müssen Sie davon Gebrauch machen, weil Sie sich nur für solche Redebeiträge interessieren, die sich den Themen Klimawandel und Nachhaltigkeit widmen. Eine solche Einschränkung lässt sich gewährleisten, indem nur solche Redebeiträge berücksichtigt werden, in denen bestimmte Schlüsselwörter vorkommen wie *Klimawandel* und *Nachhaltigkeit*.

In CQPweb können Sie mithilfe der Funktionen *Restricted Query* und *Create/edit Subcorpora* passgenau jene Texte zusammenstellen, die diesen Modellierungsparametern entsprechen. Die Funktion *Restricted Query* ermöglicht Ihnen in einem ersten Schritt, die zu untersuchenden Parteien durch Anklicken auszuwählen. Bei diesen Parteien soll nach Redebeiträgen gesucht werden, die die Wörter *Klimawandel* und *Nachhaltigkeit* enthalten. Diese Wörter müssen dafür in einen korrekt formulierten Suchbefehl überführt werden, die sogenannte CQP-Syntax:

```
[word = "Klimawandel"] | [word = "Nachhaltigkeit"]
```

Jedes Wort steht innerhalb eckiger Klammern: *Word* bezieht sich auf die Annotationsebene, auf der gesucht wird ^{III.2.2.c}. Hier

wird auf der Ebene der konkreten Wortformen gesucht, also exakt nach *Klimawandel* oder *Nachhaltigkeit*. Die ODER-Bedingung wird durch die sogenannte Pipe (|) zum Ausdruck gebracht.

Die so gefundenen Texte können Sie nun über die Funktion *Create/edit subcorpora* dauerhaft in einem Subkorpus speichern, auf das Sie bei Ihren Analysen immer wieder zugreifen können. Sie können sogar mehrere Subkorpora erstellen – das hängt letztlich davon ab, welche Analysen Sie durchführen wollen III.2.3. Nehmen Sie vergleichende Analysen vor, ist es nützlich, separate Subkorpora für jede Partei zu haben.

Erklären Sie an einem neuen Beispiel, was ein Subkorpus ist und warum Sie es bilden. Schnelltest

Auf der Webseite zum Buch finden Sie an weiterführenden Aufgaben etwa einen Screencast und eine Übung zum Erstellen von Subkorpora in CQPweb. Training

Julia Krasselt

III.2.3 Hands-on: Mit Korpora arbeiten

Ein Korpus ist wenig wert ohne ein geeignetes Set an Methoden, mit dem es analysiert werden kann. Für die Analyse von Korpora besteht ein solches Set insbesondere aus Verfahren, um Muster des Sprachgebrauchs quantitativ zu erfassen. In diesem Kapitel lernen Sie, Methoden anzuwenden, die zum essenziellen Handwerkszeug beim Forschen und Arbeiten mit Korpora gehören. So werden Sie Ihre praktischen Fähigkeiten im Umgang mit Korpora erweitern und festigen.

Lernziele **Haltung:** Sie entwickeln ein Verständnis dafür, dass Korpora mit verschiedenen, einander ergänzenden Methoden analysiert werden. **Wissen:** Sie wissen, was man in einem Korpus quantitativ messen kann. Basierend auf Beispielen lernen Sie drei zentrale Methoden der Korpusanalyse kennen. **Können:** Dieses Methodenwissen wenden Sie in Übungen an; so werden Sie sicherer im Umgang mit Korpora und Korpusanalysetools.

Aufbau Das Kapitel zeigt zunächst, was man in einem Korpus überhaupt zählen kann, um Muster im Sprachgebrauch zu entdecken ^a. Dann liegt der Fokus des Kapitels auf drei zentralen Methoden _{b–d}. Wer mit Korpora arbeitet, muss sie beherrschen.

- a Token, Types und Lemmata:
Was man in einem Korpus zählen kann
- b Distribution und Dispersion:
Verteilungen in einem Korpus erfassen
- c Konkordanzen und KWIC:
Ein Wort kommt selten allein
- d Kollokationen:
Auf den Ko-Text kommt es an

Autorin Prof. Dr. Julia Krasselt III.2.1

Thema 2.3.a Token, Types und Lemmata: Was man in einem Korpus zählen kann

Eine wichtige Methode der Korpuslinguistik ist die Messung von Frequenzen. Dabei wird erfasst, wie häufig sprachliche Einheiten in einem Korpus vorkommen. Für die Frequenzmessung sind drei Einheiten zentral: Token, Types und Lemmata.

Token sind Vorkommen oder Kopien von Einheiten, zum Beispiel von Wörtern, wie man sie auf der Sprachoberfläche zählen kann. So kann man etwa in einem Korpus auszählen, wie oft das Wort *Arzt* vorkommt und wie oft das Wort *Ärztin*. Ergebnis kann dann zum Beispiel sein, dass es 23 Token *Arzt* hat und 2 Token *Ärztin*. Dabei verkörpern die 23 Token den Type *Arzt* und die 2 Token den Type *Ärztin*. Mithilfe dieser Unterscheidung kann man in einem Korpus beispielsweise messen, wie viele verschiedene Wortbildungen (also Types) es mit *-welle* gibt (etwa *Grippewelle*, *Tsunamiwelle*) und welche davon die höchste Tokenfrequenz, also die meisten Kopien aufweist.

Korpora haben hinsichtlich der Frequenz von Types und Token eine Gemeinsamkeit: Es gibt immer nur eine kleine Anzahl hochfrequenter Types und eine sehr große Anzahl niedrigfrequenter Types. Das wird als Zipf'sche Verteilung bezeichnet, benannt nach dem Linguisten George Kingsley Zipf, der dieses Gesetz in den 1930er-Jahren aufstellte. Ein Beispiel: Die Präposition *auf* kommt in deutschsprachigen Korpora hochfrequent vor, das Substantiv *Baumwurzelnwachstum* hingegen niedrigfrequent. Die meisten Types in einem Korpus kommen sogar nur ein einziges Mal vor. Sie werden als *Hapax Legomena* bezeichnet (altgriechisch für *einmal Gesagtes*).

Der Begriff Lemma steht für die Grundform eines Wortes III.2.2.c. Mithilfe eines Lemmas können also Flexionsformen zusammengefasst werden. Bei Korpusanalysen ist das hilfreich, da bei der Frequenzmessung etwa eines Substantivs nicht alle möglichen Flexionsformen eines Wortes berücksichtigt werden müssen. Man kann so direkt nach den Token des Lemmas *Arzt* suchen, anstatt einzeln nach den Token der Types *Arztes*, *Ärzte* oder *Ärzten*.

Wie kann Frequenz angegeben werden? Frequenz wird in der Korpuslinguistik in absoluten und relativen Häufigkeiten angegeben. Die absolute Häufigkeit drückt aus, wie häufig eine Einheit im Korpus vorkommt (zum Beispiel 10-mal oder 1000-mal). Die relative Häufigkeit setzt diese absolute Häufigkeit ins Verhältnis zur Gesamtanzahl Token im Korpus. Sie wird typischerweise mit einer gut

interpretierbaren Zahl multipliziert. Das Korpustool CQPweb ^{III.2.2.c} multipliziert beispielsweise mit 1 Million und man erhält so die relative Frequenz eines Wortes pro 1 Million Wörter:

absolute Häufigkeit : Anzahl Token im Korpus / 1.000.000

Dadurch sind Aussagen wie die folgende möglich: Das Wort *Integration* hat (in einem fiktiven Korpus) eine relative Frequenz von 25.9 pro 1 Million Wörter oder anders ausgedrückt: Von einer Million Wörter sind 25.9 das Wort *Integration*. Für kleine Korpora ist es manchmal ratsam, die Referenzgröße kleiner zu wählen, etwa 100.000 oder 10.000. Die relative Frequenz ist insbesondere dann wichtig, wenn die Frequenz eines Wortes in zwei unterschiedlich großen Korpora verglichen werden soll ^{III.2.2.d}.

Schnelltest Zählen Sie die Token und Types im Satz *Alle Äpfel und alle Birnen sind frisch*.

Training Auf der Webseite zum Buch finden Sie etwa eine Übung, bei der Sie die relative Frequenz des Wortes *Integration* in zwei unterschiedlich großen Teilkorpora miteinander vergleichen.

Thema 2.3.b Distribution und Dispersion: Verteilungen in einem Korpus erfassen

Die Korpuslinguistik fragt nicht nur danach, wie häufig eine sprachliche Einheit im Korpus oder in Teilkorpora vorkommt, sondern auch, wie diese sprachliche Einheit in den Dokumenten eines Korpus verteilt ist. Anders gesagt: Die Korpuslinguistik fragt nach der Dispersion und der Distribution sprachlicher Einheiten im Korpus.

Die Dispersion beschreibt, wie gleichmäßig eine linguistische Einheit über das gesamte Korpus verteilt ist. Die Bezugsgröße sind dabei typischerweise Dokumente. Es werden Aussagen darüber getroffen, ob die untersuchte Einheit in (nahezu) allen Dokumenten des Korpus vorkommt oder nur in sehr wenigen Dokumenten. – Warum ist Dispersion relevant? Stellen Sie sich zum Beispiel vor, dass Sie typische englischsprachige Übersetzungsvarianten eines deutschen Wortes zusammenstellen wollen. Von einer typischen Variante sollten Sie erst dann sprechen, wenn sie in vielen verschiedenen Dokumenten des Korpus verwendet wird. Weniger typisch ist eine Variante, deren Vorkommen sich auf eine sehr kleine Anzahl konzentriert.

Die Distribution bezieht sich ebenfalls auf die Verteilung sprachlicher Einheiten – aber hier nicht im ganzen Korpus, sondern in Bereichen davon. Man kann sich beispielsweise anschauen, wie häufig ein Wort pro Monat in den letzten fünf Jahren gewesen ist. Dies ist dann möglich, wenn für jedes Dokument im Korpus bekannt ist, in welchem Monat es entstanden ist. Die Messung von Distribution ist eine aggregierende Methode: Das Vorkommen linguistischer Einheiten wird auf der Basis von Metadaten (hier: Monaten) zusammengefasst (aggregiert).

Die Analyse von Distribution ist für unterschiedliche Zwecke nützlich. Das ZHAW-Projekt *Wort des Jahres* fragt etwa, welche Wörter im Vergleich zum Vorjahr deutlich häufiger in den Medien verwendet werden Perrin et al., 2020. Wörter, die im Vergleich zu den Vorjahren im aktuellen Jahr deutlich frequenter sind, stellen potenzielle Kandidaten für das Wort des Jahres in der Schweiz dar. Visuell sind solche Analysen Verlaufskurven: Die folgende Abbildung zeigt, dass das Wort *Schutzstatus* bis Anfang 2022 in Zeitungen der Deutschschweiz kaum verwendet wird und dann aufgrund des Krieges in der Ukraine an Relevanz gewinnt Abb. 4.

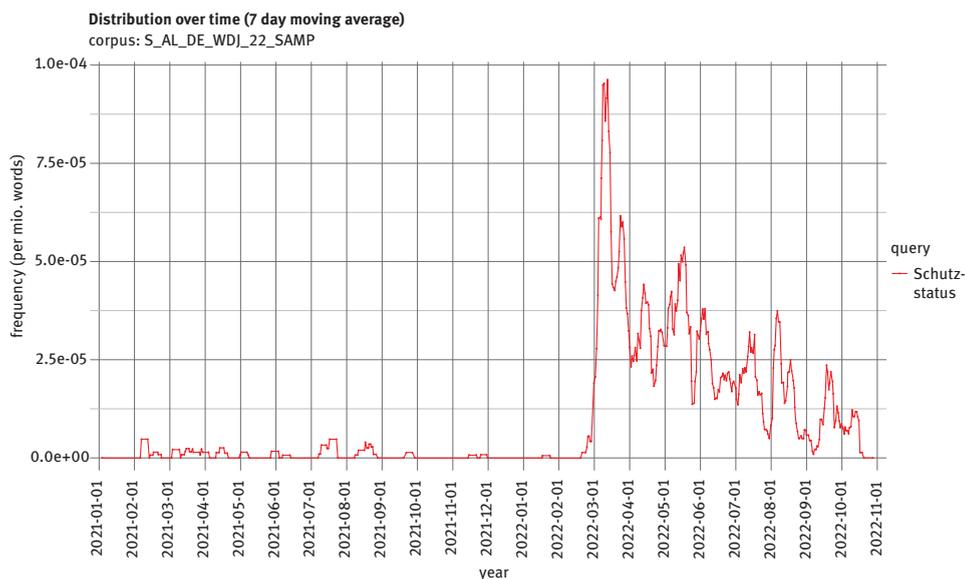


Abb. 4: Distribution des Wortes *Schutzstatus* in Deutschschweizer Medien.

<https://tools.linguistik.zhaw.ch/rsconnect/workbench/>

- Schnelltest Erklären Sie den Unterschied zwischen Distribution und Dispersion.
- Training Auf der Webseite zum Buch finden Sie etwa eine Übung, bei der Sie die Distribution der Wörter *Integration* und *Inklusion* in Deutschschweizer Medien seit 2015 herausfinden.

Thema 2.3.c Konkordanz und KWIC: Ein Wort kommt selten allein

Neben den aggregierenden Methoden ^{III.2.3.b} ist in der Korpuslinguistik ebenso die Analyse einzelner Textstellen zentral. Dies geschieht über sogenannte Konkordanzen. Dabei werden in einem Korpusanalysetool alle Fundstellen eines Wortes oder einer Wortfolge untereinander dargestellt, samt dem direkt vorausgehenden linken Ko-Text ^{II.2.2.c} und dem direkt nachfolgenden rechten Ko-Text. Eine andere Bezeichnung für eine solche Darstellung ist KWIC, *Keyword in Context*. Die folgende Abbildung zeigt eine solche Ansicht für die Suche *Adjektiv + Demokratie* ^{III.2.2.c} im Tool CQPweb ^{Abb. 5}:

	Solution 1 to 50 Page	1/28
dienvielfalt , die Medienfreiheit sind Säulen der	direkten Demokratie	. Deswegen haben Sie erst kürzlich beispielsweise
s und balances „ - , die Sie vorhin als Segen der	direkten Demokratie	beschworen haben und für welche Sie sogar 80
en können , wer sie hier vertritt , und in unserer	direkten Demokratie	abstimmen können . Eine direkte Demokratie , d
direkten Demokratie abstimmen können . Eine	direkte Demokratie	, die stolz auf sich ist - zu Recht stolz auf sich
tens um die wichtige Frage , wie wir zu unserer	direkten Demokratie	Sorge tragen wollen - und das müssen wir unbe
en . Wenn wir dies nicht tun , schwächen wir die	direkte Demokratie	, und das will die BDP auf keinen Fall . Wir bitter
ament , führte er weiter aus , müsse ein Ort der	lebendigen Demokratie	bleiben , stets inspiriert durch den Respekt vor
ern es geht um ein gesamtheitliches System der	rechtsstaatlichen Demokratie	, in dem sowohl Fisch wie Vogel ihren Platz und
leutlich , dass das konstruktive Referendum die	direkte Demokratie	überfordert . Eine Gesetzesvorlage gleicht scho

Abb. 5: Konkordanzen für die Suche *Adjektiv + Demokratie* in CQPweb.

<https://swiss-al.linguistik.zhaw.ch/CQPweb/parldebatten/>

Erste korpuslinguistische Analysetools wurden früher als *Concordancer* bezeichnet, weil sie genau über diese eine Funktion verfügten: Fundstellen in Form von Konkordanzen darstellen. Auch heute bieten alle Korpusanalysetools die Möglichkeit, Konkordanzen anzuzeigen. Bei einigen Tools kann man die Größe des dargestellten linken und rechten Kontexts einstellen, bei anderen Tools ist dieses Kontextfenster vorgegeben.

Konkordanzen sind also – anders als die Messung von Frequenz, Distribution bzw. Dispersion – keine Form der Aggregation III.2.3.a. Stattdessen ermöglichen sie es, den Gebrauch eines Wortes in seinem unmittelbaren Kontext anzusehen und sehr nah an den Dokumenten zu sein, aus denen das Korpus besteht. Konkordanzen können einerseits zu den quantitativen Methoden gezählt werden, weil sie die Analyse aller Belegstellen ermöglichen. Andererseits ermöglichen sie aber eben auch eine qualitative Analyse auf der Ebene einzelner Textstellen.

Konkordanzen werden mit dem Ziel analysiert, Muster in der Verwendung eines Wortes zu entdecken. Durch das alphabetische Sortieren nach einem der Kontextwörter stehen ähnliche Belegstellen direkt untereinander. Die folgende Abbildung zeigt, dass *Demokratie* musterhaft mit anderen sogenannten Hochwertwörtern (d. h. Wörtern, die in einer Gesellschaft akzeptierte Grundwerte bezeichnen) wie *Freiheit* kombiniert wird. Erst durch das Sortieren wird dieses Sprachgebrauchsmuster sichtbar Abb. 6.

Etwa ein Viertel von ihnen ist im Einsatz für **Freiheit** und **Demokratie** gestorben . Und genau das begründet ihre historische Le
 Völker im arabischen Raum in ihrem Kampf für **Freiheit** , **Demokratie** und sozialen Wandel zu unterstützen . Gerade wenn wir
 rderlichen Glaubwürdigkeit und mit Mut für **Freiheit** und **Demokratie** einzustehen . Dabei sind besonders auch jene Kräfte zu
 von Menschen demonstrieren seit Wochen für **Freiheit** , **Demokratie** und sozialen Wandel und gegen die Diktatoren ihrer Län
 gliedstaaten die Bürgerinnen und Bürger in **Freiheit** und **Demokratie** über ihren Willen abgestimmt haben , dieser Organisati
 nmen , denn die Menschen brauchen neben **Freiheit** und **Demokratie** auch ein Auskommen für ihre Familien . Die SP-Fraktion
 Staaten diktieren lassen , deren Oberhäuptern **Freiheit** , **Demokratie** , freie Meinungsäußerung , Pressefreiheit , Chancenge
 und damit auch für die Werte der Schweizer **Freiheit** und **Demokratie** , die in ihrer Geschichte nie so bedroht waren wie durch
 unveräußerlicher Rechte des Menschen sowie **Freiheit** , **Demokratie** , Gleichheit und Rechtsstaatlichkeit als universelle Wert

Abb. 6: Sortierte Konkordanz für die Suche *Demokratie* in CQPweb.

<https://swiss-al.linguistik.zhaw.ch/CQPweb/parldebatten/>

Was spricht dafür, in der Analyse großer Korpora die Konkordanz- Schnelltest
 analyse zu nutzen?

Auf der Webseite zum Buch finden Sie weiterführende Aufgaben, Training
 zum Beispiel eine Übung, bei der Sie Konkordanzen eines Wortes
 sortieren und Muster identifizieren.

Thema 2.3.d Kollokationen:

Auf den Ko-Text kommt es an

Sieht man sich in einem Medienkorpus für 2020, also das erste Jahr der COVID-19-Pandemie, die Konkordanzen III.2.3.a für das Wort *Ab-*

stand an, dann wird man sehr viele Belegstellen entdecken, in denen auch die Wörter *Maßnahme*, *Meter*, *halten*, *zueinander* oder *vorgeschrieben* vorkommen. Diese Wörter kommen sogar so häufig im Ko-Text von *Abstand* vor, dass man nicht mehr von Zufall sprechen kann. All diese Wörter sind im Jahr 2020 sogenannte Kollokationen des Wortes *Abstand*. Unter einer Kollokation versteht man in der Korpuslinguistik ein Wort, das überzufällig häufig im Ko-Text eines anderen Wortes vorkommt.

Die Beobachtung, dass bestimmte Wörter häufiger miteinander vorkommen als andere, ist deutlich älter als die Korpuslinguistik. Der Begriff *Kollokation* wurde zum ersten Mal vom US-amerikanischen Linguisten John Rupert Firth verwendet und gehört inzwischen zu den zentralen Konzepten in der Korpuslinguistik. Firth hat nicht nur festgestellt, dass gewisse Wörter sehr häufig gemeinsam verwendet werden, sondern hat daraus auch Aussagen über die Bedeutung von Wörtern abgeleitet. Firth prägte den Satz: „You shall know a word by the company it keeps“ Firth, 1957, 179. Um herauszufinden, was ein Wort bedeutet, muss man sich also anschauen, in welchem Ko-Text es verwendet wird. Daraus folgt auch, dass Wörter verschiedene Bedeutungen haben können, wenn sie in unterschiedlichen Zusammenhängen verwendet werden.

In einem Korpus können nun für jedes Wort große Mengen an Ko-Text-Informationen gesammelt werden. Es kann beispielsweise gezählt werden, welche Wörter wie häufig innerhalb eines vordefinierten Fensters gemeinsam mit einem gesuchten Wort vorkommen. Bezogen auf das Eingangsbeispiel kann man auszählen, welche Wörter links und rechts vom Wort *Abstand* stehen. Anschließend werden diese beobachteten Häufigkeiten verglichen mit den erwarteten Häufigkeiten. Tritt dann das Wort *zueinander* deutlich häufiger im Kontext von *Abstand* auf, als es die Häufigkeit dieser beiden Wörter im Korpus erwarten lässt, dann ist *zueinander* eine Kollokation von *Abstand*. In Korpusanalysetools stehen dafür verschiedene statistische Tests und Maßzahlen zur Verfügung.

Um Kollokationen für Aussagen über die Bedeutung eines Wortes zu verwenden, ist eine Kategorisierung und Interpretation durch Forschende notwendig. Anders gesagt: Eine Liste mit Kollokationen, etwa zum Wort *Abstand*, ist noch kein fertiges korpuslinguistisches Analyseergebnis! Beispielsweise könnten Kollokationen nach grammatischen, pragmatischen oder thematischen Gesichtspunkten kategorisiert werden. Für das Wort *Abstand* würde man

für das Corona-Jahr 2020 beispielsweise feststellen, dass es geprägt ist durch Wörter aus den Bedeutungsfeldern *Hygiene* und *Distanz*: Die Bedeutung des Wortes *Abstand* bezieht sich also auf die physische Distanz von Personen zum Zweck der Ansteckungsvermeidung.

Erklären Sie, warum Kollokationen Aussagen über die Bedeutung eines Wortes ermöglichen. Schnelltest

Auf der Webseite zum Buch finden Sie weitere Übungen. Zum Beispiel berechnen und interpretieren Sie Kollokationen eines Wortes mit dem Log-Likelihood-Test. Training

Julia Krasselt, Peter Stücheli-Herlach, Davide Garassino,
Elsa Liste Lamas und Maren Runte

III.2.4 Praxisprobleme lösen: Korpuslinguistik im Beruf

In den Berufsfeldern der Angewandten Linguistik wird es Ihre Aufgabe sein, kommunikative Probleme zu lösen. In diesem Kapitel lernen Sie, wie Korpora und korpuslinguistische Methoden Ihnen dabei helfen, als *Reflective Practitioner* zu agieren, indem Sie vom tatsächlichen Sprachgebrauch ausgehen, wie er in großen Korpora erfasst ist. Im Fokus des Kapitels stehen die Berufsfelder Journalismus und Organisationskommunikation, Übersetzen und Dolmetschen sowie sprachliche Integration.

Lernziele **Haltung:** Sie entwickeln ein Verständnis dafür, dass professionelles Arbeiten mit Sprache voraussetzt, die Standards und Muster des Sprachgebrauchs zu kennen, und der Einsatz von Korpora dazu dient, diese Standards und Muster zu identifizieren. **Wissen:** Sie lernen, in welchen Praxisfeldern die Korpuslinguistik eingesetzt wird und welche Probleme sie dort löst. **Können:** Sie wenden Ihr bisher erworbenes Wissen in Praxisbeispielen an.

Aufbau Das Kapitel zeigt die Praxisrelevanz der Korpuslinguistik ^a und verdeutlicht dies anhand von drei ausgewählten Praxisfeldern ^{b–d}.

- a Praxisrelevanz von Korpuslinguistik:
Die Sprache gelingender Kommunikation
- b Organisationskommunikation und Journalismus:
Öffentliche Diskurse mit Korpora erforschen
- c Übersetzen und Dolmetschen:
Sprachmittlung im Zeitalter von Korpora
- d Sprache und Integration:
Mit Korpora lernen für DaF-/DaZ-Kurse

Autor:innen Prof. Dr. Peter Stücheli-Herlach ist Professor für Organisationskommunikation und Öffentlichkeit am ZHAW Institut für Angewandte Medienwissenschaften und Mitbegründer des Digital Discourse Lab und der Swiss-AL-Korpora an der ZHAW.

<https://www.zhaw.ch/de/ueber-uns/person/stue/>

Dr. Davide Garassino ist Dozent am ZHAW Institut für Übersetzen und Dolmetschen. Er hat langjährige Lehr- und Forschungserfahrung in den Bereichen Pragmatik, Prosodie und Korpuslinguistik. Unter anderem hat er an der Erstellung und der sprachvergleichenden Analyse des multilingualen Korpus ICOCP mitgewirkt.

<https://www.zhaw.ch/de/ueber-uns/person/gars/>

Dr. Elsa Liste Lamas ist wissenschaftliche Mitarbeiterin am ZHAW Institut of Language Competence. Sie hat Forschungs- und Lehrerfahrung in den Bereichen Fremd-/Zweitspracherwerb sowie Korpuslinguistik. Unter anderem hat sie am Aufbau des Parallelkorpus *PaGeS* mitgewirkt und ist für die ILC-Lernendenkorpora *LeKoBe* und *LeKoD-CH* verantwortlich.

<https://www.zhaw.ch/de/ueber-uns/person/list/>

Dr. Maren Runte ^{1.1.2} ist Dozentin am ZHAW Institute of Language Competence. Sie hat langjährige Forschungserfahrung in den Bereichen Lexikografie und Korpuslinguistik. In der Forschung zu Wörterbuchbenutzung zeigt sie, wie sich Auswertungen von großen Mengen an Sprachdaten benutzer:innenadäquat darstellen lassen.

<https://www.zhaw.ch/de/ueber-uns/person/runm/>

Thema 2.4.a Praxisrelevanz von Korpuslinguistik: Die Sprache gelingender Kommunikation

Professionelles Arbeiten mit Sprache setzt voraus, dass wir Standards und Muster des Sprachgebrauchs kennen und unser Handeln an diesem Wissen ausrichten. Es geht also darum, kompetent erheben und beurteilen zu können, was Menschen mit Sprache im Alltag machen – und was die Sprache im Alltag mit den Menschen macht. Durch konkreten Gebrauch kommt die Sprache zur Geltung und Wirkung.

Etwas mit Sprache zu tun heißt, die eigene Äußerung so zu gestalten, dass sie zur gestellten Aufgabe, zur Situation und zu den Adressat:innen passt. Die geeigneten Mittel wählen die Kommunizierenden möglichst zielführend aus, indem sie abwägen unter den Varianten in der gewählten Sprache und Varietät ^{1.2.2}. Eigene Äußerungen sollen also auf sprachliche, kognitive und soziale Kontexte – auf Diskurse – Bezug nehmen: Sprachprofis setzen Zeichen, die für andere sinnvoll sind.

Professionelles Sprachhandeln in Diskurse einzubetten, ist Diskurshandeln. Diskurshandeln evident erklären und gezielt verbes-

sern zu können, ist Diskurskompetenz. Die Korpuslinguistik, ihre Methoden und Erkenntnisse unterstützen kompetentes Diskurshandeln besonders in der digitalisierten Arbeitswelt. Korpuslinguistik – und die darauf bauende Diskurslinguistik – ermöglichen Sprachprofis, größere Sammlungen von Daten zu spezifischen Diskursen zu beurteilen und nutzen, sie zu erstellen und weiterzuentwickeln. Diese Kompetenzen stärken ihre Handlungs- und Konkurrenzfähigkeit.

In Organisationskommunikation und Journalismus dient der Einsatz von Sprachkorpora besonders der Umfeldanalyse, etwa der Recherche über Themenfelder und Wertegemeinschaften, welche die Öffentlichkeit prägen. Die Aussagenproduktion wird so anschlussfähig an relevante Wissensbestände und Deutungsmuster.

In der Sprachmittlung dienen Korpora der Klärung terminologischer Fragen und phraseologischer Standards. Sprachprofis entwickeln und beurteilen dann zum Beispiel Übersetzungsäquivalente zielsicherer. Gemeinsamkeiten und Unterschiede verschiedener Domänen und Kulturgemeinschaften erschließen sich so präziser.

Für sprachliche Integration ist die Arbeit mit Lernendenkorpora besonders zielführend. Aus solchen Korpora lässt sich Wissen über Spracherwerb oder konkrete Varietäten von Lernenden gewinnen. Zudem liefern Korpusanalysen Lern- und Übungsbeispiele, die zu den Lernzielen, den Lernenden und deren situativen Anwendungen passen.

Sprachprofis stehen also im Dienst von Sprach- und Diskursgemeinschaften. Aus diesen Gemeinschaften entsteht eine Nachfrage nach professionellen Leistungen – und eine Bereitschaft, diese Leistungen letztlich auch zu entschädigen. Routinen und Bedürfnisse dieser Gemeinschaften erschließen sich über deren Sprachgebrauch – kurz: In der Sprache zeigt sich, wie sie ticken. Korpus- und Diskurslinguistik liefern die geeigneten Werkzeuge, dies zu erkennen. Wer sie einsetzt, begründet Entscheidungen präziser. Und beweist überzeugend, dass es auf sie oder auf ihn ankommt.

- Schnelltest Erklären Sie die Bedeutung von Korpuslinguistik für Ihr Berufsfeld – in Ihren eigenen Worten.
- Training Auf der Webseite zum Buch finden Sie zum Beispiel einen Werbespot, dessen Risiken und Chancen in interkultureller Kommunikation Sie einschätzen.

Thema 2.4.b Organisationskommunikation und Journalismus: Öffentliche Diskurse mit Korpora erforschen

Organisationskommunikation und Journalismus sind Berufsfelder, die öffentlichen Diskurs in der Gesellschaft herstellen, was im Wesentlichen durch spezifischen Sprachgebrauch geschieht. Profis in diesen Feldern arbeiten also selbst mit der Sprache – und reflektieren zugleich, auf welche spezifische Art sie und andere sich der Sprache bedienen, um Wirkung zu entfalten.

Muster des Sprachgebrauchs, wie sie die Korpuslinguistik untersucht, sind deshalb in Organisationskommunikation und Journalismus entscheidend. Es gibt Organisationsdiskurse, in denen sich Erwartungen von Menschen aneinander musterhaft manifestieren und die entsprechend Aushandlungen von Angebot und Nachfrage, von Preisen und Werten prägen. Es gibt journalistische Diskurse, in denen sich musterhaft manifestiert, was jeweils in einer Zeitperiode für mehrere Funktionssysteme der Gesellschaft gleichermaßen relevant ist.

Die Korpuslinguistik gibt Sprachprofis die Werkzeuge an die Hand, um solche Muster zu identifizieren und kritisch zu reflektieren. Das unterstützt begründetes und zielführendes berufliches Wirken, denn:

- Dieser Muster bedienen sie sich selber, etwa mit Schlagworten und Redensarten, aber auch Deutungsmustern. Korpuslinguistische Diskursanalyse ermöglicht die Selbstreflexion und unterstützt die wirkungsorientierte Textproduktion.
- Diese Muster möchten sie mitprägen, etwa durch gute Storys, die von anderen aufgegriffen werden sollen, oder durch Issues-Management und Corporate Messaging für Organisationen. Korpuslinguistische Diskursanalyse unterstützt Recherchen und ermöglicht es Organisationen, zu erkennen, wie sie und ihre Anliegen in der Öffentlichkeit wahrgenommen und bewertet werden.
- Diese Muster müssen sie erkennen, um Normen fürs Sprachhandeln reflektieren zu können, etwa in Gender-, Rassismus- und Populismusdebatten. Korpuslinguistische Diskursanalyse erlaubt es, eigene Positionen zu klären und die Perspektiven anderer besser zu verstehen.

Die Anwendung der Korpuslinguistik für die Diskursanalyse in diesen Berufsfeldern ist vielfältig möglich Dreesen & Stücheli-Herlach, 2019

- Korpora sind Modelle ^{III.2.2.a} von öffentlichen Diskursen auf konkreten Kommunikationsfeldern mit ihren Akteur:innen und Interessen (wie Energiepolitik, Nachhaltigkeit oder Automobilität).
- Frequenz-, Distributions- und Dispersionsanalysen ^{III.2.3} beleuchten thematische Strukturen öffentlicher Diskurse (wie die Rede von *erneuerbaren Energien* oder *nachhaltigem Bauen*).
- Konkordanz- und Kollokationsanalysen ^{III.2.3} ergründen spezifische Formen musterhaften Schreibens in der Öffentlichkeit, wie es sich bezüglich der Rollen einzelner Akteur:innen oder in Public Stories ausprägen kann, etwa über selbstbezügliche oder gemeinnützige Rollen, über gesellschaftlichen Fortschritt oder Konflikte.

Schnelltest	Warum sollen sich Profis öffentlicher Kommunikation für Sprachdaten-Korpora interessieren?
Training	Auf der Webseite finden Sie zum Beispiel eine Übung, bei der Sie den Gebrauch des Schlagworts <i>Freedom Day</i> in journalistischen und Organisationsdiskursen untersuchen.

Thema 2.4.c Übersetzen und Dolmetschen: Sprachmittlung im Zeitalter von Korpora

Die alltägliche Arbeit von Übersetzer:innen und Dolmetscher:innen profitiert stark von der Verfügbarkeit empirischer Daten. Dies zeigt sich beispielsweise in der Nutzung von Werkzeugen wie den CAT-Tools ^{II.2.2.a} und maschinellen Übersetzungen, die auf enormen Datenbanken von Quell- und Zieltexten beruhen.

Auch der Beitrag linguistischer Korpora ist für die Übersetzungs- und Dolmetschpraxis von großer Bedeutung. Anhand von Korpora, die aus übersetzten oder gedolmetschten Materialien bestehen, wie etwa EPIC, dem European Parliament Interpretation Corpus ^{Russo et al., 2012}, können Übersetzer:innen und Dolmetscher:innen beispielsweise Strategien beobachten, die andere Expert:innen bei bestimmten Sprachproblemen anwenden. Eine Stärke von Korpora liegt darin, dass sie im Vergleich zu Werkzeugen wie Wörterbüchern zahlreiche Informationen über den Ko-Text sowie die Frequenz von Wörtern und größeren sprachlichen Einheiten bieten. Dies kann Übersetzer:innen dabei helfen, terminologische Probleme zu lösen und für den jeweiligen Kontext geeignete Übersetzungsäquivalente zu finden.

Außerdem können Übersetzer:innen dank der Nutzer:innenfreundlichkeit von Tools wie AntConc Anthony, 2023 vergleichbare oder parallele Mikrokorpora für spezifische Textsorten erstellen, die sie zur Beantwortung ebenso spezifischer Fragen verwenden können: Wie lässt sich beispielsweise *bone [re]growth* auf einer deutschen Webseite über kosmetische Zahnmedizin optimal übersetzen Bernardini, 2022?

Ein wichtiger Aspekt bei Übersetzungen ist die mögliche Interferenz der Quellsprache, also die Übertragung sprachlicher Muster von der Originalsprache in die Zielsprache. Parallelkorpora III.2.2.b erlauben es, solche Interferenzen in verschiedenen Textsorten direkt zu beobachten. Darüber hinaus ermöglichen vergleichbare Korpora III.2.2.b es, sich mit den lexikalischen und stilistischen Merkmalen der Originaltexte – wie zum Beispiel den üblichsten Formulierungen bestimmter Textsorten – vertraut zu machen.

Dank der Möglichkeit, Informationen in authentischen Kontexten aufzufinden, sind Korpora ein optimales Instrument für Dolmetscher:innen, die passende Fachterminologie und Phraseologie sowie auch fachspezifisches Wissen zu erwerben, etwa im Bereich der internationalen Politik.

In Anbetracht der wichtigen Rolle des Dolmetschens, insbesondere in institutionellen Kontexten, könnte man erwarten, dass zahlreiche Korpora zur Verfügung stehen. Aufgrund der Schwierigkeiten bei der Erstellung von Korpora der gesprochenen Sprache und der Entwicklung von Transkriptionssystemen waren jedoch lange nur wenige Korpora verfügbar, die zudem oft sehr klein und auf wenige Textsorten beschränkt waren. Erfreulicherweise ändert sich jedoch auch in diesem Bereich die Situation rasch: So existieren heute etwa das EPIC s. o. und das EPTIC, das *European Parliament Translation and Interpreting Corpus*, das mehrere europäische Forschungsgruppen gemeinsam entwickelt haben Ferraresi & Bernardini, 2019.

Welche Vorteile bieten parallele Korpora für Übersetzer:innen und Dolmetscher:innen?

Schnelltest

Auf der Webseite zum Buch üben Sie weiter, etwa mit einem parallelen Korpus von deutschen und englischen Untertiteln.

Training

Thema 2.4.d Sprache und Integration: Mit Korpora lernen für DaF-/DaZ-Kurse

Auch in der Sprachlern- und -lehrforschung sowie der Fremd- und Zweitsprachendidaktik spielen Korpora eine zunehmend wichtige Rolle. Hierbei muss unterschieden werden, wer welches Korpus mit welcher Absicht erstellt und verwendet. Der nachfolgende Überblick orientiert sich an Schwerpunkten der Korpuslinguistik in der Lehre von Deutsch als Fremd- und Zweitsprache (DaF/DaZ).

In der Forschung zum DaF/DaZ-Erwerb bilden Lernendenkorpora III.2.2.b eine wichtige Datengrundlage für die Beschreibung von Varietäten I.2.2 von Lernenden sowie von Verläufen des Spracherwerbs I.2.1. In solchen Korpora sammeln Forschende mündliche und schriftliche Texte von Deutschlernenden und annotieren sie in Bezug auf lernendensprachliche Merkmale wie Normverstöße.

Dank der Arbeit mit Lernendenkorpora können Forschende über die Beschreibung kleiner Lernendengruppen hinausgehen. Sie führen umfassende und systematische Analysen durch, die ein besseres Verständnis der Sprachstandsentwicklung sowie des Spracherwerbsprozesses ermöglichen. Je nach Korpus können auch Lernendenvarietäten verschiedener Erstsprachen (L1) und Kompetenzstufen verglichen werden. Zum Beispiel können Forschende untersuchen, wie sich die Verbstellung mit zunehmendem Kompetenzniveau entwickelt und ob sich Unterschiede je nach L1 feststellen lassen. Die Erkenntnisse, die durch Lernendenkorpora gewonnen werden, können als Grundlage für die Entwicklung von didaktischen Materialien dienen, etwa von Übungen zu spezifischen Schwierigkeiten der Lernenden mit dem Satzbau.

Neben Lernendenkorpora werden auch allgemeine Korpora III.2.2.b für die Entwicklung von Lern-, Lehr- und Prüfungsmaterialien eingesetzt. Der Mehrwert liegt darin, dass dann die Beispiele und Übungen auf authentischem Sprachgebrauch basieren. Eine konkrete Anwendung ist die Berücksichtigung von Häufigkeitsklassen bei der Erarbeitung von Wortschatzlisten: Häufig gebrauchte Lexik kann so früher angeboten werden als selten(er) gebrauchte Lexik.

DaF-/DaZ-Lehrende können Korpora nutzen, um den Lernenden spezifische Aspekte des Sprachgebrauchs nahezubringen. Dabei können Lernende anhand von authentischen Sprachdaten sprachliche Zusammenhänge bis hin zu Sprachwandelphänomenen

erfassen, etwa den Kasuswechsel bei der Präposition *wegen*. Die hierbei abgefragten Korpora sollen möglichst repräsentativ sein für die jeweilige Sprache – und idealerweise öffentlich zugänglich.

Bislang wenig praktiziert wird der Einsatz von Korpora in der Ausbildung von Deutschlehrenden. Sinnvoll wäre es, wenn bereits in der Ausbildung die Arbeit mit allgemeinen Korpora wie dem Deutschen Referenzkorpus (DeReKo) Leibniz-Institut für Deutsche Sprache, 2022, dem Digitalen Wörterbuch der deutschen Sprache (DWDS) Berlin-Brandenburgische Akademie der Wissenschaften, o. J. oder mit Lernendenkorpora wie dem fehlerannotierten Lernerkorpus des Deutschen (Falko) Lüdeling et al., 2008 oder dem Schweizer DaF-/DaZ-Lernendenkorpus (LeKoD-CH) systematisch integriert würde.

Nennen Sie je drei Vorteile von Korpora für Sprachlehrende und -lernende. Schnelltest

Auf der Webseite zum Buch lernen Sie zum Beispiel in einer Übung, mit welchem Kasus die Präposition *während* typischerweise verwendet wird. Training