

KORPORA in DaF und DaZ: Theorie und Praxis

Carolina Flinz

1 Einleitung

Die moderne Korpuslinguistik wird zum ersten Mal in einer systematischen und unterrichtspraktisch verwertbaren Weise für den Englischunterricht von Mukherjee (Mukherjee 2002: 28) dargestellt:

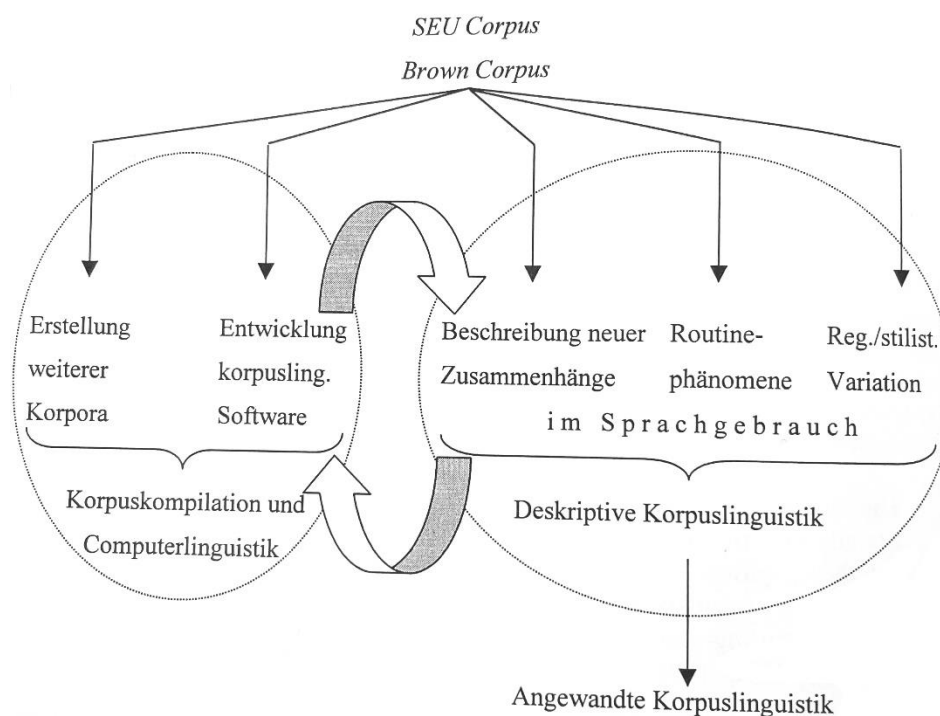


Abb. 1: Die moderne Korpuslinguistik nach Mukherjee (2002: 28)

Dort wird sie als „deskriptive Korpuslinguistik“ beschrieben, die in Wechselbeziehung mit der Computerlinguistik steht und deren Befunde Anwendungsbereiche auch in der Didaktik und der Methodik des Fremdsprachenunterrichts haben. Als wichtigste Entwicklungslinien des Zeitraums 1961–2001 werden die Beschreibung neuer Zusammenhänge, Routinephänomene sowie regionale und stilistische Variationen (auch auf einer diachronischen Ebene) im Sprachgebrauch vermerkt.

Für die Fächer *Deutsch als Fremdsprache* (DaF) und *Deutsch als Zweitsprache* (DaZ)¹ liegt ein solcher Versuch noch nicht vor, auch wenn schon 2007 Fandrych/Tschirner eine spezifische Einführung für DaF als „dringendes Desideratum“ bewerten, da DaF, dessen Ziel die Erforschung fremdsprachlichen Lehrens und Lernens ist, eine „deutliche Affinität“ zu empirischen Methoden hat (Fandrych/Tschirner 2007: 195). 2009 betonen Lüdeling/Walter erneut, dass die meisten Verfahren und Szenarien bisher für das Englische als Fremdsprache entwickelt worden sind; sie begründen dabei ihre mögliche Anwendung für DaF (2009: 3), denn sie könnten ihrer Ansicht nach problemlos übertragen werden. 2013 definiert Wallner (2013) den Eingang der Korpuslinguistik in die DaF-Didaktik noch als zögerlich; 2019 und 2020 wird auf die Skepsis der Lehrenden eingegangen und es wird für eine stärkere Benutzung von Korpora im DaF-Unterricht plädiert (vgl. Flinz/Katelhön 2019 und Flinz 2020).

Dieser chronologische Unterschied zwischen dem Eingang der Korpuslinguistik in *Englisch as a Foreign Language* (EFL) und in DAF ist damit verbunden, dass die Korpuslinguistik als angewandte Disziplin, auch dank des Kontextualismus von Firth, in der englischsprachigen Forschung eine längere Tradition² und erst später in Deutschland Fuß gefasst hat. Die ersten germanistischen Einführungen in die Korpuslinguistik stammen nämlich aus dem Jahr 2006 (vgl. Lemnitzer/Zinsmeister 2006 und Scherer 2006)³. Alle bezeugen, dass die Disziplin nicht nur wichtig für die Beschreibung des Sprachgebrauchs ist, sondern auch für die Fremdsprachendidaktik und den Fremdsprachenunterricht (vgl. Lemnitzer/Zinsmeister 2015: 33–34).

Trotz der chronologischen Diskrepanz zwischen den zwei Sprachräumen und den erwiesenen Vorteilen, die auch empirisch belegt worden sind (vgl. Abschnitt 2), ist

¹ Im Folgenden werden DaF und DaZ nicht mehr getrennt. Alle Hinweise auf DaF können auch auf DaZ bezogen werden.

² Vgl. u.a. McEnery/Xiao/Tono 2006; Mukherjee 2002, 2008; Römer 2008. Erwähnenswert ist auch das COBUILD Projekt (vgl. Sinclair 1987), das einen starken Einfluss auf die Lernerlexikographie, Lernergrammatikographie und Lehrwerkentwicklung nahm.

³ 2010 wurde die zweite Ausgabe von Lemnitzer/Zinsmeister (2010) veröffentlicht und 2012 haben Perkuhn/Keibel/Kupietz das Taschenbuch *Korpuslinguistik* publiziert. 2014 erschien dann die zweite Ausgabe der Einführung von Scherer, und 2015 die dritte Ausgabe von Lemnitzer/Zinsmeister. 2019 wurde die Einführung von Hirschmann veröffentlicht.

der induktive Zugang zu sprachlichen Phänomenen, auch als *Data Driven Learning* (DDL) bekannt, weiterhin noch ziemlich unbekannt innerhalb der Fremdsprachendidaktik (Boulton 2020: XV): „DDL [Data-driven Learning] work in applied linguistics remains relatively unknown in educational circles“. Die Benutzung von Korpora in der Fremdsprachendidaktik wurde damals als revolutionär bewertet (vgl. Rundell/Stock 1992), da sie zeigen konnte, wie Sprache in authentischen Kommunikationssituationen gebraucht wird. Die Ergebnisse der Analysen konnten dann in Wörterbücher, Grammatiken, Lehrbücher, Kursmaterialien, Lehrpläne einfließen (vgl. Boulton/Tyne 2013: 97), auch wenn im Gegensatz zum angelsächsischen Raum in Deutschland die Auswirkungen bei der Lehrbuchgestaltung geringer und daher noch stärker erwünscht sind (vgl. Lemnitzer/Zinsmeister 2015: 192).

Trotz des gestiegenen Interesses ist die damalige Feststellung von Römer (2009: 92) noch heute gültig: Die Visibilität des DDLs muss noch mehr steigen, so dass sowohl Lehrende als auch Lernende davon profitieren können. Sie sollten stärker mit den üblichen Aktivitäten im Unterricht verbunden werden (vgl. Mukherjee 2006) und als etwas Alltägliches präsentiert werden (vgl. Boulton/Tyne 2013: 98), so dass die Skepsis gegenüber diesem Ansatz gemindert werden kann.

Mit dieser Themenausgabe möchten wir einen kleinen Schritt in diese Richtung gehen. Nach einem Überblick über die Vorteile des *Data Driven Learning* für das Fach *Deutsch als Fremdsprache*, in dem auch die damit verbundenen Hürden angesprochen werden (Abschnitt 2), werden unterschiedliche Typen von Korpora der deutschen Sprache präsentiert (Abschnitt 3). Im Abschnitt 4 werden mögliche Methoden und Ansätze der Korpuslinguistik im DaF-Unterricht besprochen, und im Abschnitt 5 werden die einzelnen Beiträge der vorliegenden Themenausgabe vorgestellt.

2 Data Driven Learning und Deutsch als Fremdsprache

Data Driven Learning wurden im Laufe der Zeit mehrere Bedeutungen beigegeben (vgl. Boulton/Tyne 2013: 98). In diesem Aufsatz wird von einer erweiterten Perspektive des DDLs ausgegangen: Das DDL ist mit seiner Benutzung von Tools und Methoden der Korpuslinguistik nicht nur für fremdsprachendidaktische Ziele interessant, sondern es werden auch pädagogische Zwecke mit diesem Ansatz verfolgt (vgl. Gilquin/Granger 2010: 359). Es handelt sich beim DDL also um eine Methode, die sich auf eine Verbindung von unterschiedlichen Ansätzen und Theorien stützt (Flowerdew 2015):

- *Noticing*, d.h. das sprachliche Material bewusst wahrnehmen (vgl. Portmann-Tselikas 2001);
- Konstruktivismus: Lernende sind die ProtagonistInnen und Kompetenzen werden dynamisch erworben;
- sozio-kulturelle Interaktion: Die Interaktion zwischen Lehrenden und Lernenden ist zentral, da Lehrende als Regisseure den Fremdsprachenerwerb leiten, so dass *scaffolding* und Ko-Konstruktion von Erkenntnissen unterstützt werden;
- Fokus auf Lernstile und kognitive Prozesse.

Lehrende können Korpora benutzen, um mit authentischen sprachlichen Daten im Unterricht induktiv (Regeln aus Daten) sowie auch deduktiv (Exemplifizierung von Regeln auf der Basis von Daten) zu arbeiten: So können zum Beispiel Form und Funktion (bzw. Funktionen, da viele Elemente der deutschen Sprache polyfunktional sind) in Verbindung gebracht werden (vgl. Flinz 2020). Die ausgewählte Arbeitsweise – induktiv oder deduktiv – hängt von den Bedürfnissen der Lernenden (vgl. Johansson 2009) und den Lernzielen ab.

Lernende können Korpora in unterschiedlicher Weise benutzen und bewegen sich zwischen zwei entgegengesetzten Polen: Kontrollierte Aufgaben und Übungen, die von der Lehrperson vermittelt werden, auf der einen Seite und exploratives und autonomes Recherchieren auf der anderen Seite (vgl. der Unterschied zwischen „soft“ DDL und „hard“ DDL in Gabrielatos 2005: 17). Lernende können nämlich in Korpora vollständig autonom recherchieren und Korpora auch selbst erstellen: Sie sind *Forscher* (Johns 1988: 14) oder *Sherlock Holmes* (Johns 1997: 101), die die Sprache autonom erkunden.

Die DDL-Didaktik hat die Rolle, den Lehrenden und Lernenden sprachliche Belege zu liefern, so dass induktiv Informationen zu Strukturen, Kontexten, Gebrauch und Funktionen entnommen werden können; diese sind nicht immer in den benutzten Lehrbüchern enthalten (vgl. Corino 2019: 273).

Viele Studien und Aufsätze heben wiederholt die Vorteile des DDL hervor. Boulton/Tyne fassen diese folgendermaßen zusammen:

DDL is alleged to enhance cognitive and metacognitive skills, increase sensitivity to authentic language use, provide an interactive approach to constructivist discovery learning, foster motivation especially through individualisation, promote reusable and transferable skills, favour autonomy for life-long learning, and correspond largely to current theories of second language acquisition (Boulton/Tyne 2013: 99).

Kognitive und metakognitive Kompetenzen werden gefördert, Sensibilität für authentischen Sprachgebrauch wird geschult, exploratives Lernen wird intensiviert, Motivation und Autonomie werden gesteigert.

Auch Zweitspracherwerbstheorien befürworten diese Methode (vgl. u.a. das *lexical priming* in Hoey 2005 und das *mental corpus* in Taylor 2012). Bei den Vorteilen berufen sie sich auch auf schon obengenannte Thesen der Fremdsprachendidaktik, wie den Vorteil des Gebrauchs von authentischen Sprachdaten (vgl. Widdowson 2000); das Benutzen von Inputdaten, die das Rohmaterial für das Gehirn liefern (vgl. Fandrych/Tschirner 2007: 196); das Benutzen von Inputfaktoren, die den Spracherwerb beeinflussen (Häufigkeit, Salienz, Komplexität, Kontext, vgl. Fandrych/Tschirner 2007: 198–199); den Konstruktivismus (vgl. Widdowson 2000), das bewusste Lernen (vgl. Fandrych/Tschirner 2007: 196), den lernerzentrierten Ansatz (vgl. Kirschner/ Sweller/Clark 2006), den aufgabenbasierte Ansatz (vgl. Swan 2005)⁴, die Benutzung von kontrastiven Analysen, da der Vergleich eine wichtige Grundlage des fremdsprachlichen Lernens darstellt (vgl. Fandrych/Tschirner 2007: 196), die Sprachaufmerksamkeit (vgl. Portmann-Tselikas 2001) etc.

Auch empirische Studien belegten weitere Vorteile⁵, wie z.B., dass die Benutzung von Korpora die schriftliche Sprachproduktion fördert (vgl. Boulton/Tyne 2013: 100). Vorteile aus einer Mittelzeitstudie wurden auch bestätigt (vgl. Boulton 2011); Langzeitstudien hingegen sind noch eine Rarität (vgl. Mukherjee 2006). Lerner-typen und Sprachenniveaus wurden bislang ebenfalls nur teilweise in den Studien berücksichtigt. Varley (2009) and Boulton (2010) zeigen jedoch, dass DDL auch als Alternative für Lernende, die mit anderen Methoden wenig Erfolg hatten, benutzt werden kann. Corino (2019: 275) fügt hinzu, dass zum einen dank der immer stärker benutzerfreundlichen Tools, die saliente Elemente graphisch hervorheben, visuelle Lerntypen begünstigt werden und dass zum anderen durch die autonome Arbeit mit den Daten auch haptische Lerntypen profitieren können⁶. In diesem Aufsatz wird davon ausgegangen, dass nicht nur visuelle und haptische Lerntypen mit dem DDL begünstigt werden, sondern auch auditive Lerntypen, insbesondere, wenn mit Korpora der gesprochenen Sprache gearbeitet wird, einen Zugewinn ha-

⁴ Für eine Diskussion dazu vgl. Boulton 2009.

⁵ Boulton (2020) zitiert 378 Veröffentlichungen, die auf der Basis von empirischen Studien die Vorteile von DDL für den Spracherwerb hervorheben.

⁶ Nach Corino (2019) sind Studien zur Verbindung zwischen DDL, Konstruktivismus und Lernstilen stark erwünscht.

ben und auch kommunikative Lerntypen nicht benachteiligt sind, da mit der Gruppenarbeit und der Aufgabenbasierung Kommunikation untereinander und im Plenum veranlasst wird.

Trotz all der erwiesenen Vorteile wird mit Ausnahme einiger Kreise und Projekte DDL immer noch wenig im Fremdsprachenunterricht wahrgenommen, wenn nicht vollständig ignoriert (vgl. Corino 2019: 272), insbesondere im Schulbereich. Auf universitärer Ebene gibt es Versuche einzelner DozentInnen. Eine Begründung hierfür kann die fehlende Aus- und Weiterbildung der Lehrenden sein (auch wenn Korpuslinguistik in vielen Studiengängen präsent ist), die Skepsis der Lehrenden, die die Tools sowie die metakognitive bzw. kognitive Anstrengung, die für ihre Verwendung gebraucht wird, als eine Hürde sehen, der Aufbau einer *corpus literacy* (vgl. Mukherijee 2002), der Zeitfaktor für die Planung der Aktivitäten, die Erstellung der Materialien sowie das Vertrautmachen der Lernenden mit diesen (vgl. Corino/Onesti 2019). Corino (2019) hebt in Anlehnung an Römer (2006) in Bezug auf die Lehrpersonen besonders die mangelnde Aus- bzw. Weiterbildung hervor, so dass die Verbreitung dieser Methode nicht als Teil des Lehrauftrags verstanden wird. Das Fehlen einer neuen Lehrenden-Generation, die 2007 von Braun (2007: 308) gefordert wurde, kann noch heute festgestellt werden.

3 Die deutsche Korpuslandschaft

„Ein Korpus ist eine Sammlung schriftlicher oder gesprochener Äußerungen in einer oder mehreren Sprachen in digitaler Form“ (Lemnitzer/Zinsmeister 2015: 13). Was ein Korpus von einer Sammlung authentischer Daten unterscheidet, die schon seit langer Zeit in der Fremdsprachendidaktik verwendet werden (Lüdeling/Walter 2009: 1), ist die „digitale Form“ und die Benutzung von „korpuslinguistischen Tools“ anstatt manueller Verfahren für ihre Analyse. Ein Korpus besteht meistens aber nicht nur aus den primären digitalen Daten (die *Primärdaten* sind Texte oder Äußerungen, die das Korpus konstituieren), sondern auch aus den Informationen über sie (die *Metadaten*) und die *linguistischen Annotationen*, d.h. Zusatzinformationen, die unterschiedliche sprachliche Ebenen betreffen können.

Die deutsche Korpuslandschaft ist vielfältig: Man kann unterschiedliche Korpus Typen finden, die unterschiedliche Metadaten und Annotationsmöglichkeiten bereitstellen. Auflistungen von Korpora können den gängigen Einführungen zur Korpuslinguistik (vgl. Lemnitzer/Zinsmeister 2015: 142–156) oder der Clarin-Umgebung entnommen werden (<https://www.clarin.eu/resource-families>). In diesem Abschnitt soll ein Überblick gegeben werden. Als Orientierung wird die Klassifizierung von Lemnitzer/Zinsmeister (2015: 137) dienen, die Korpora auf der Basis unterschiedlicher Kriterien, die sowohl Designkriterien (*sampling*-Kriterien) als auch Kriterien

zur Korpusaufbereitung (Annotation) beinhalten, unterscheidet: Funktionalität, Sprachenauswahl, Medium, Größe, Annotation, Persistenz, Sprachbezug, Verfügbarkeit (vgl. Abbildung 2):

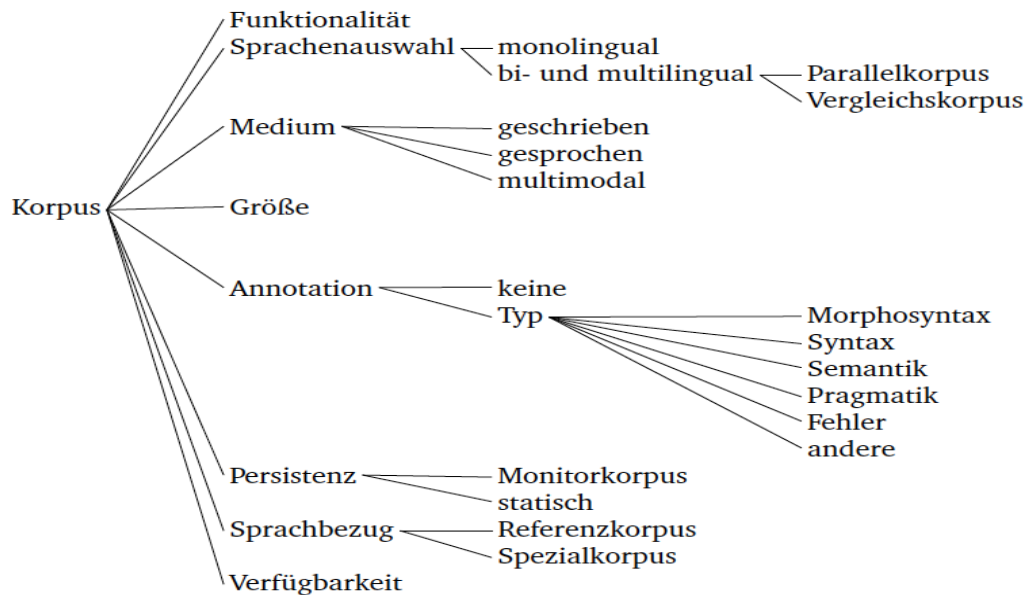


Abb. 2: Korpustypologie nach Lemnitzer/Zinsmeister (2015: 137)

Das erste Kriterium ist die Funktionalität, die besagt, zu welchem Zweck ein Korpus erstellt wurde. Sie ist ein sehr wichtiges Kriterium, da sie auch einen Einfluss auf die weiteren Kriterien hat. Die ursprüngliche Funktionalität eines Korpus kann im Laufe der Zeit variieren, und viele Korpora sind auch polyfunktional.

Das zweite Kriterium ist die Sprachenauswahl: Sie betrifft die Sprache der Dokumente des Korpus. Ein Korpus kann eine, zwei oder mehrere Sprachen enthalten: Es kann also *monolingual*, *bilingual* oder *multilingual* sein. Im Falle von bilingualen und multilingualen Korpora wird zwischen *Parallelkorpora* (bzw. Übersetzungen) und *Vergleichskorpora* (Texte in mehreren Sprachen, die zu vergleichbaren Diskursbereichen erfasst worden sind, aber keine Übersetzungen sind) unterschieden.

Das dritte Kriterium ist das Medium. Es bezieht sich auf die Primärdaten. Man unterscheidet zwischen *Korpora der gesprochenen Sprache*, in denen man dazu auch die Transkripte rechnet (vgl. Sinclair 1996), *Korpora der geschriebenen Sprache*, in denen schriftliche Texte enthalten sind, auch wenn sie wie eine geschriebene Rede konzeptionell mündlich wären (vgl. Koch/Oesterreicher 1994) und *multimodale Korpora*, in denen unterschiedliche Medien enthalten sind (Video, Audio etc.). Der letzten Gruppe sind auch Gestik-Korpora zuzurechnen.

Das vierte Kriterium ist die Annotation. Sie kann vorhanden sein oder nicht. Im ersten Fall kann es sich um eine automatische, semiautomatische oder manuelle Annotation handeln, die sich auf unterschiedliche Sprachebenen beziehen kann (Morphosyntax, Syntax, Semantik, Pragmatik, Fehler, Textstruktur, Informationsstruktur, Phonetik/Prosodie, Gestik etc.). Im zweiten Fall ist das Korpus nicht annotiert und die Rohdaten können explorativ untersucht werden.

Das fünfte Kriterium ist die Größe eines Korpus. Die Größe kann stark variieren. Heute können Korpora die Größe von mehreren Milliarden Textwörtern erreichen, wie beim *Deutschen Referenzkorpus* (DeReKo); die Größe kann weiterhin ständig wachsen (*Monitorkorpora*). Die Größe ist nicht nur bei quantitativen Untersuchungen (vgl. Abschnitt 4) relevant, da valide statistische Aussagen nur möglich sind, wenn eine große Menge an Daten vorhanden ist, sondern auch für kleinere korpusgestützte Studien, die sich auf seltene Phänomene konzentrieren. Diese können nur in großen Datenmengen entdeckt werden. Die Größe eines Korpus sollte aber nicht das Kriterium für jede Forschungsfrage sein, denn wenn man *Spezialkorpora* untersucht, kann auch eine begrenzte Zahl an Textwörtern ausreichen (vgl. Flinz/Perkuhn 2018; Flinz 2021). Die Größe eines Korpus ist stark mit der Forschungsfrage verbunden.

Das sechste Kriterium ist die Persistenz. Sie ist mit der Aktualisierung der Primärdaten verbunden. Ein Korpus kann statisch sein und somit nicht mehr aktualisiert, erweitert und verändert werden (*statisches Korpus*); wird das Korpus hingegen erweitert und modifiziert, ist es ein *Monitorkorpus*⁷.

Das siebte Kriterium ist der Bezug zum Untersuchungsgegenstand. Es ist mit den Eigenschaften des zu repräsentierenden Gegenstandes verbunden. Auf einer Skala wären ganz oben die *Referenzkorpora*, die den Gegenstand äußerst gut und komplett darstellen, wie zum Beispiel eine natürliche Sprache in einem bestimmten Zeitraum. Referenzkorpora können auch als Kontrollkorpora genutzt werden, wenn bestimmte Eigenschaften eines *Spezialkorpus* – ein Korpus, das sich auf ein bestimmtes Thema, einen bestimmten Sachverhalt, eine bestimmte Fachsprache, einen bestimmten Autor etc. konzentriert – vergleichend untersucht werden. Bei einigen Korpusmanagement-Tools, wie zum Beispiel *Sketch Engine*, sind Referenzkorpora genau zu diesem Zweck in der Funktionalität *Keywords* enthalten: Man kann das eigene Spezialkorpus (dort als *Fokuskorpus* dargestellt) mit dem als Default eingestellten Referenzkorpus vergleichen und die *Keywords* auf dieser Basis extrahieren.

⁷ Perkuhn/Keibel/Kupietz (2012: 51) unterscheiden hinzu zwischen dynamischem Korpus und Monitorkorpus.

Das achte Kriterium ist die Verfügbarkeit, die mit der Benutzung der Daten zusammenhängt. Neben Korpora, die durch Schnittstellen frei untersucht werden können oder die sogar heruntergeladen werden können, gibt es Korpora, bei denen man sich registrieren oder einen Lizenzvertrag abschließen muss.

Es folgt eine kleine Auflistung der gängigsten Korpora der deutschen Sprache. Es handelt sich um eine kleine Auswahl, die sich auf folgende Kriterien stützt: freie Verfügbarkeit, Einfluss auf korpuslinguistische Untersuchungen und Verwendbarkeit, auch durch Aufsätze bestätigt, für das Fach DaF.

3.1 Das Deutsche Referenzkorpus (DeReKo)

Tab. 1: Eigenschaften von DeReKo

Funktionalität	Linguistische Forschung
Sprachenauswahl	Monolingual: hauptsächlich Deutsch, aber bei einigen Korpora sind auch weitere Sprachen enthalten (vgl. <i>WP-Wikipedia Korpora</i> , die als Vergleichskorpora klassifiziert werden können).
Medium	Geschriebene Sprache. Die Medialität ist in einigen Subkorpora unterschiedlich, da dort auch konzeptuell mündliche Korpora enthalten sind (u.a. <i>Dortmunder Chat Korpus</i>).
Größe	50,6 Milliarden Textwörter (Stand 02.02.2021).
Annotation	Alle Korpora sind mehrfach annotiert, Zugang zu Annotationen hängt vom ausgewählten Recherchesystem ab.
Persistenz	Auch als Monitorkorpus nutzbar
Untersuchungsgegenstand	Referenzkorpus. Einige Subkorpora sind Spezialkorpora. Beliebige Teilmengen als virtuelles Korpus definierbar.
Verfügbarkeit	Zum Download nur bestimmte Korpora nach Absprache, frei recherchierbar nach kostenloser Registrierung über die Systeme Cosmas II und KorAp.

Das *Deutsche Referenzkorpus DeReKo (DeReKo)* ist mit seinen 50,6 Milliarden Textwörtern (Stand 02.02.2021) „die weltweit größte Sammlung deutschsprachiger Korpora“⁸. Es enthält belletristische, wissenschaftliche und populärwissenschaftliche Texte, Zeitungstexte sowie weitere Textarten, die in Hinblick auf Umfang, Variabilität, Qualität und Aktualität ausgewählt werden.

Die (urheberrechtlich geschützten)⁹ geschriebenen deutschsprachigen Texte entstammen der Gegenwart, einschließlich der jüngeren Vergangenheit. Das nutzungsrechtlich abgesicherte Material wird kontinuierlich weiterentwickelt, ergänzt und unterliegt stetigem Wachstum: ca. 3.1 Millionen Textwörter pro Jahr (Kupietz/Lüngen/Kamocki/Witt 2018: 4353).

DeReKo beinhaltet außer Archiven, die mehrere Korpusstypen enthalten, auch unterschiedliche Korpora, die einzeln ausgewählt werden können, darunter auch historische Korpora (wie *das Goethe Korpus* und *das Grimm Korpus*), literarische Korpora (*biographische Literatur*), spezifische Zeitungskorpora (u.a. *österreichische Zeitungskorpora*, *Bonner Zeitungskorpus*) Spezialkorpora (*Fußball-Liveticker*, *Fachsprachenkorpora*, *Handbuchkorpora*), CMC-Korpora (*Dortmunder Chatkorpus 2.2*), vergleichbare Korpora (*die Wikipedia Korpora*).

Die Korpora sind mit dem Korpusrecherche- und Analysesystem COSMAS II (*Corpus Search, Management and Analysis System*) durchsuch- und analysierbar¹⁰. Das System bietet den NutzerInnen die Möglichkeit, virtuelle Korpora zu erstellen.

Die Suchergebnisse (vgl. die Suche nach *Pandemie* im Archiv W im vordefiniertem Korpus W-öffentlich, die 1944 Treffer ergibt) können nach Jahrzehnt, Jahr, Textsorte, Thema etc. gefiltert werden (Abbildung 3):

⁸ Vgl. <https://www1.ids-mannheim.de/kl/projekte/korpora/>

⁹ Vgl. u.a. W-Archiv der geschriebenen Sprache (auch W2, W3, W4, W-Übrig); PARONYM - Archiv des Paronym-Projekts des IDS; WP-Wikipedia Artikel (auch WPE und WP-FS), TAGGED-Corpora (TAGGED-C; TAGGED-C2; TAGGED-T; TAGGED-T2; TAGGED-M); HIST-Archiv der historischen Korpora; UMB-Archiv des Projekts *Sprachliche Umbrüche des 20. Jahrhunderts*; GFDS-Kartei der Gesellschaft für deutsche Sprache; WK-PH-Archiv der phasengegliederten Wendekorpora.

¹⁰ Alle Leistungen der Korpusrecherchewerkzeuge *COSMAS I und II* sollen in der Zukunft von *KorAP* übernommen werden, das bereits in beta-Version angeboten wird.

Aktuelles Archiv: W - Archiv der geschriebenen Sprache **Aktuelles Korpus:** W-öffentlich - alle öffentlichen Korpora des Archivs W (mit Neuaquisitionen) [1]

Aktuelle Suchanfrage: Pandemie **Referenz:** Deutsches Referenzkorpus (DeReKo-2020-I)

Treffer: 7.907 **Aktive Treffer:**

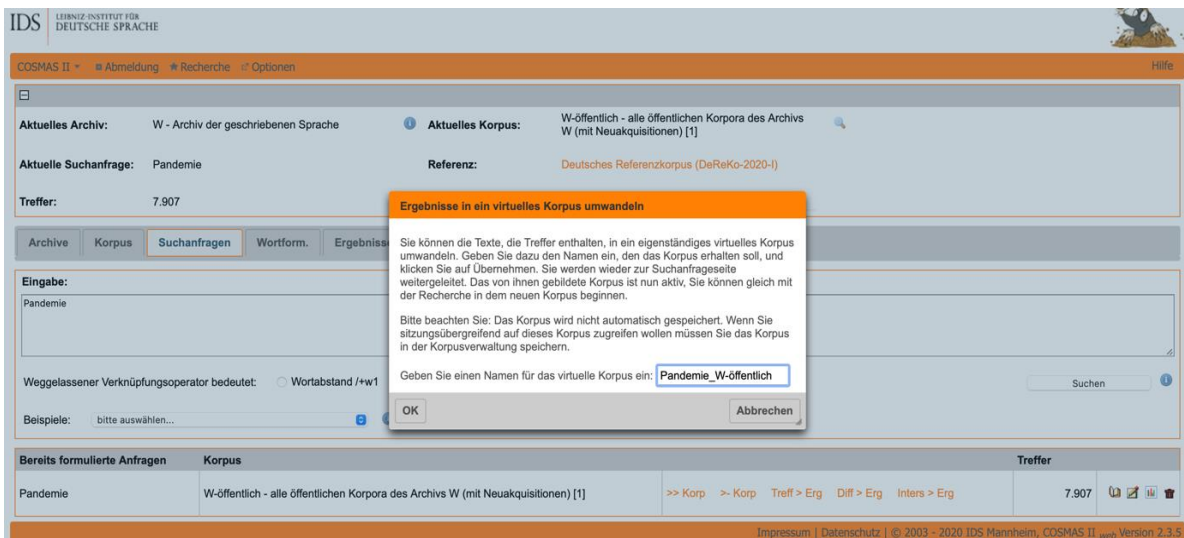
Archiv | Korpus | Such. | Wortform. | **Ergebnisse** | Kook. | KWIC | Volltext | Export

Quellenansicht
 ✓ Korpusansicht
 Dokumentansicht
 Ansicht nach Ländern
 Ansicht nach Textsorten
 Ansicht nach Themen
 Ansicht nach Jahrzehnt
 Ansicht nach Jahr
 Ansicht nach Monat
 Ansicht nach Tag
 Ansicht vor/seit Zeitpunkt
 Ansicht nach Mehrwortspanne
 Ansicht nach Wort-Types
 Ansicht nach Wortendungen

is	Korpus
997	A97 St. Galler Tagblatt 1997
017	B17 Berliner Zeitung 2017
019	B19 Berliner Zeitung 2019
013	BRZ13 Braunschweiger Zeitung 2013
019	CT19 ct 2019
014	FLT14 Falter 2014
012	FOC12 FOCUS 2012
017	FOC17 FOCUS 2017
008	HAZ08 Hannoversche Allgemeine 2008
016	HAZ16 Hannoversche Allgemeine 2016
1	1 2007 HMP07 Hamburger Morgenpost 2007
1	1 2012 HMP12 Hamburger Morgenpost 2012
1	1 2013 2013 HMP13 Hamburger Morgenpost 2013
1	1 2015 2015 HMP15 Hamburger Morgenpost 2015
1	1 1997 1997 I97 Tiroler Tageszeitung 1997
1	1 2001 2001 L01 Berliner Morgenpost 2001
1	1 2002 2002 L02 Berliner Morgenpost 2002
1	1 2003 2003 L03 Berliner Morgenpost 2003
1	1 2013 2013 LTB13 Luxemburger Tageblatt 2013
1	1 2016 2016 LTB16 Luxemburger Tageblatt 2016
7.907	4.672 1953 2019 556 Korpora

Abb. 3: Sortiermöglichkeiten der Suchergebnisse

Die Ergebnisse der Recherche können auch als virtuelles Korpus gespeichert werden, so dass der Zugriff darauf für weitere nachfolgende Untersuchungen gewährleistet ist. Wenn man unter den bereits formulierten Anfragen auf das erste orange Korpus klickt (>> **Korp**), erscheint eine Maske mit dem Hinweis *Ergebnisse in ein virtuelles Korpus umwandeln*; der Name des Korpus soll dabei ausgesucht und gespeichert werden (Abbildung 4):



IDS LEIBNIZ-INSTITUT FÜR DEUTSCHE SPRACHE

COSMAS II - Abmeldung - Recherche - Optionen

Aktuelles Archiv: W - Archiv der geschriebenen Sprache **Aktuelles Korpus:** W-öffentlich - alle öffentlichen Korpora des Archivs W (mit Neuaquisitionen) [1]

Aktuelle Suchanfrage: Pandemie **Referenz:** Deutsches Referenzkorpus (DeReKo-2020-I)

Treffer: 7.907

Archiv | Korpus | Suchanfragen | Wortform. | **Ergebnisse**

Eingabe: Pandemie

Weggelassener Verknüpfungoperator bedeutet: Wortabstand /+w1

Beispiele: bitte auswählen...

Ergebnisse in ein virtuelles Korpus umwandeln

Sie können die Texte, die Treffer enthalten, in ein eigenständiges virtuelles Korpus umwandeln. Geben Sie dazu den Namen ein, den das Korpus erhalten soll, und klicken Sie auf Übernehmen. Sie werden wieder zur Suchanfrageseite weitergeleitet. Das von Ihnen gebildete Korpus ist nun aktiv, Sie können gleich mit der Recherche in dem neuen Korpus beginnen.

Bitte beachten Sie: Das Korpus wird nicht automatisch gespeichert. Wenn Sie sitzungsübergreifend auf dieses Korpus zugreifen wollen müssen Sie das Korpus in der Korpusverwaltung speichern.

Geben Sie einen Namen für das virtuelle Korpus ein:

OK Abbrechen

Bereits formulierte Anfragen **Korpus** **Treffer**

Pandemie	W-öffentlich - alle öffentlichen Korpora des Archivs W (mit Neuaquisitionen) [1]	>> Korp >> Korp Treff > Erg Diff > Erg Inters > Erg	7.907
----------	--	--	-------

Impressum | Datenschutz | © 2003 - 2020 IDS Mannheim, COSMAS II web Version 2.3.5

Abb. 4: Speicherung der Ergebnisse als virtuelles Korpus

DeReKo ist für sprachwissenschaftliche Zwecke, nicht für die Didaktik konzipiert, kann aber durchaus dazu benutzt werden (vgl. u.a. Nolting/Radke 2019), auch wenn eine gewisse Methodenkompetenz und Einarbeitungszeit für die effiziente Benutzung von DeReKo notwendig ist (vgl. Steyer 2008: 189). Die Möglichkeit, virtuelle

Korpora zu erstellen, die repräsentativ oder auf spezielle Aufgabenstellungen kalibriert sind, ist für die DaF-Didaktik von besonderer Bedeutung, da die Lernenden mit kleineren zugeschnittenen Korpora arbeiten können und die Gefahr, dass sie nicht zu den erwünschten Ergebnissen kommen, weil sie von der Fülle von Daten überflutet werden, gemindert wird.

Da die Auswahl des Korpus und die Formulierung der Suchanfrage für die Aussagekraft der Ergebnisse maßgeblich ist, bietet die Infrastruktur viele Möglichkeiten, nimmt den Nutzer/die Nutzerin allerdings auch in die Pflicht, sich mit beidem bewusst auseinanderzusetzen. Insbesondere auch bei der Suchanfrage ist zu entscheiden, welches sprachliche Objekt „erfasst“ werden soll – ob es sich um eine Wortform handelt, ob Groß-/Kleinschreibung relevant ist, ob ein gesamtes Paradigma gemeinsam betrachtet wird oder ob sogar Wortbildung mitberücksichtigt werden soll. Diese im Vergleich zu anderen Angeboten anfänglich als unangenehm empfundene Bürde macht sich nach kurzer Zeit bezahlt für die reflektierende Auseinandersetzung mit „Sprache“, insbesondere mit ihrem korpuslinguistischen Zugang.

Für die Untersuchung von Beziehungen zwischen Wörtern (zunächst primär syntagmatisch, aber auch paradigmatisch) können automatisch berechnete Kookkurrenzen in DeReKo mit zwei Tools erstellt werden:

1. Die statistische Kookkurrenzanalyse, die in COSMAS II implementiert ist; sie ermittelt signifikante Wortkombinationen, nicht nur als Bigramme, sondern auch als rekurrente syntagmatische Strukturen:

Aktuelles Archiv:	W - Archiv der geschriebenen Sprache	Aktuelles Korpus:	Pandemie_W-öffentlich [2]
Aktuelle Suchanfrage:	Pandemie	Referenz:	Deutsches Referenzkorpus (DeReKo-2020-I)
Treffer:	7.907	Aktive Treffer:	

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

Einstellungen	Kookkurrenzen
---------------	----------------------

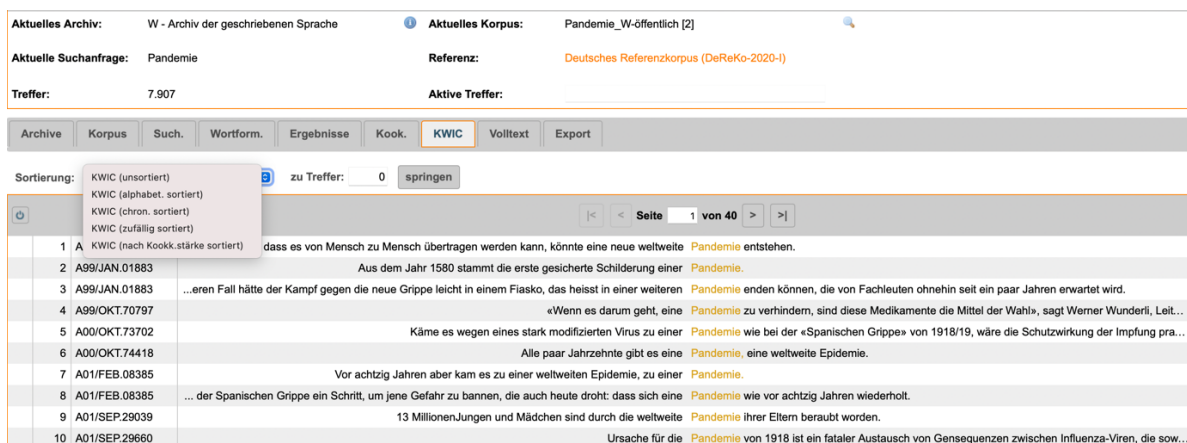
#	LLR	kumul.	Häufig	Kookkurrenzen	syntagmatische Muster
1	2445	202	202	Schweinegrippe	78% die Schweinegrippe [zur] Pandemie erklärt
2	1887	364	162	Epidemie	61% einer weltweiten Epidemie [... einer] Pandemie
3	1533	520	156	weltweiten	73% vor einer weltweiten [...] Pandemie
4	1449	561	41	Grippe Spanische	60% die eine Pandemie [wie die] Spanische Grippe
		668	107	Grippe	52% Grippe [...] die eine] Pandemie
5	1405	800	132	weltweite	66% eine weltweite [...] Pandemie
6	1395	936	136	auslösen	99% eine Pandemie [...] auslösen könnte
7	1335	1128	192	Fälle	99% im Im Falle einer Pandemie
8	1184	1139	11	H1N1 WEITERLEITUNG	100% WEITERLEITUNG Pandemie H1N1 2009/10
		1186	47	H1N1	59% Pandemie [...] H1N1 2009 2009/10 ...
9	1183	1320	134	Ausbruch	95% den Ausbruch der einer Pandemie
10	1110	1408	88	Virus	60% das Virus das ... eine Pandemie

Abb. 5: Ergebnisse der Kookkurrenzanalyse von *Pandemie*

Keywords in Context (KWICs) und Korpusstellen werden in Signifikanzclustern zusammengestellt, wobei die Parameter (u.a. Kookkurrenzspanne, Korpusbasis) individuell festgelegt werden können. Die Suche nach *Pandemie* hat u.a. ergeben,

- dass das Adjektiv *weltweite* häufig (132 Okkurrenzen) mit *Pandemie* rekurriert und 66% im syntagmatischen Muster *eine weltweite Pandemie*;
- dass das Verb *auslösen* 136 Mal mit *Pandemie* im Muster *eine Pandemie [...] auslösen könnte* vorkommt;
- dass das Nomen *Falle* 192 Mal im Muster *im Falle einer Pandemie* (99%) benutzt wird.
- Lernende können somit über die attributive Funktion von Adjektiven reflektieren, sowie über Kollokationen und Mehrwortverbindungen.

Anschließend können die KWICs oder die betreffenden Textstellen nach unterschiedlichen Kriterien sortiert werden (unsortiert, alphabetisch sortiert, chronologisch sortiert, zufällig sortiert, nach Kookkurrenzstärke sortiert, vgl. Abbildung 6):



The screenshot shows the KWIC search interface. At the top, the search parameters are: 'Aktuelles Archiv: W - Archiv der geschriebenen Sprache', 'Aktuelles Korpus: Pandemie_W-öffentlich [2]', 'Aktuelle Suchanfrage: Pandemie', 'Referenz: Deutsches Referenzkorpus (DeReKo-2020-I)', and 'Treffer: 7.907'. Below this is a navigation bar with tabs for 'Archive', 'Korpus', 'Such.', 'Wortform.', 'Ergebnisse', 'Kook.', 'KWIC' (selected), 'Volltext', and 'Export'. A dropdown menu for 'Sortierung:' is open, showing options: 'KWIC (unsortiert)', 'KWIC (alphabet. sortiert)', 'KWIC (chron. sortiert)', 'KWIC (zufällig sortiert)', and 'KWIC (nach Kookk.stärke sortiert)'. The main results table shows 10 entries, with the first entry selected. The table columns are 'KWIC (nach Kookk.stärke sortiert)' and the corresponding text snippet. The text snippets contain the word 'Pandemie' highlighted in orange.

KWIC (nach Kookk.stärke sortiert)	Text
1 A	dass es von Mensch zu Mensch übertragen werden kann, könnte eine neue weltweite Pandemie entstehen.
2 A99/JAN.01883	Aus dem Jahr 1580 stammt die erste gesicherte Schilderung einer Pandemie .
3 A99/JAN.01883	...eren Fall hätte der Kampf gegen die neue Grippe leicht in einem Fiasko, das heisst in einer weiteren Pandemie enden können, die von Fachleuten ohnehin seit ein paar Jahren erwartet wird.
4 A99/OKT.70797	«Wenn es darum geht, eine Pandemie zu verhindern, sind diese Medikamente die Mittel der Wahl», sagt Werner Wunderli, Leit...
5 A00/OKT.73702	Käme es wegen eines stark modifizierten Virus zu einer Pandemie wie bei der «Spanischen Grippe» von 1918/19, wäre die Schutzwirkung der Impfung prä...
6 A00/OKT.74418	Alle paar Jahrzehnte gibt es eine Pandemie , eine weltweite Epidemie.
7 A01/FEB.08385	Vor achtzig Jahren aber kam es zu einer weltweiten Epidemie, zu einer Pandemie .
8 A01/FEB.08385	... der Spanischen Grippe ein Schritt, um jene Gefahr zu bannen, die auch heute droht: dass sich eine Pandemie wie vor achtzig Jahren wiederholt.
9 A01/SEP.29039	13 MillionenJungen und Mädchen sind durch die weltweite Pandemie ihrer Eltern beraubt worden.
10 A01/SEP.29660	Ursache für die Pandemie von 1918 ist ein fataler Austausch von Gensequenzen zwischen Influenza-Viren, die sow...

Abb. 6: Sortierungsmöglichkeiten der KWICs

In der dargestellten Abbildung 6 sind die Ergebnisse unsortiert geordnet. Die Ergebnisse können anschließend exportiert werden.

2. Die *Kookkurrenzdatenbank* (CCDB¹¹), die automatisch berechnete Kookkurrenzprofile zu 220.000 Wörtern visualisiert (vgl. Abbildung 7, die die Ergebnisse zu *Pandemie* darstellt).

¹¹ Vgl. <http://corpora.ids-mannheim.de/ccdb/>

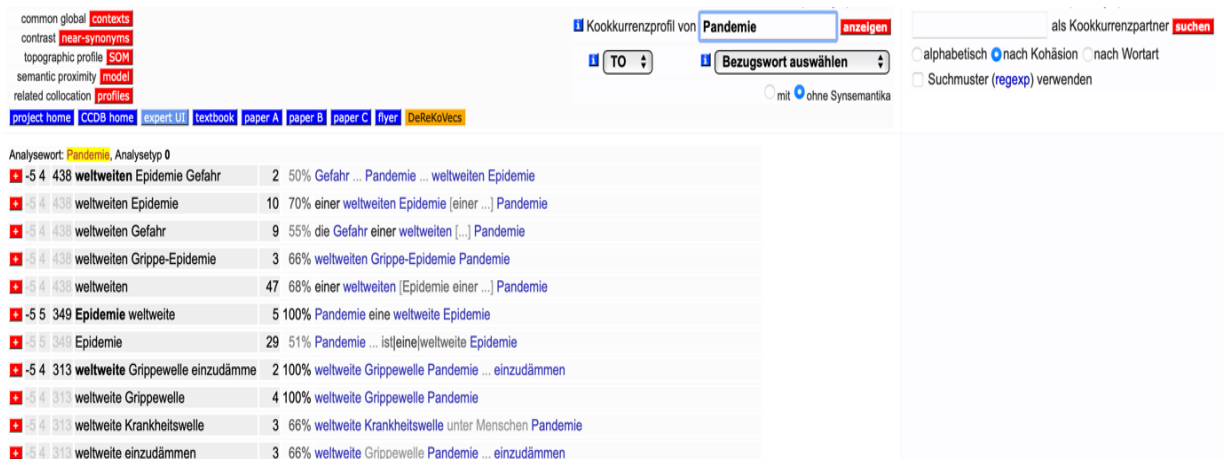


Abb. 7: Ergebnisse der Kookkurrenzanalyse von *Pandemie* in der CCDB

Die Kookkurrenzanalyse hat die bevorzugten Partnerwörter von *Pandemie* identifiziert. Diese Muster können ein wichtiger Einstiegspunkt bei der Besprechung sprachlicher Phänomene auf unterschiedlicher Sprachebenen sein. Diese Suche könnte zum Beispiel Anlass für die Reflektionen über mögliche Komposita des Wortes *Pandemie* sein, sowie die Ähnlichkeiten/Unterschiede im Gebrauch von *Epidemie*.

Die CCDB bietet weiterhin die Möglichkeit, im Gebrauch ähnliche Wörter anzuzeigen (related profiles), sowie die Gebrauchsaspekte als Self-Organizing Map (SOM), auch mit einem Vergleichswort (CNS), zu visualisieren. Da die Datengrundlage für die CCDB allerdings 2007 eingefroren wurde, sind neuere Tendenzen in der Sprache damit nicht aufzuspüren. Dafür steht eine alternative Plattform zur Verfügung, die sich im frühen Experimentierstadium befindet (<http://corpora.ids-mannheim.de/openlab/derekovecs/>).

3.2 Das DWDS-Korpus

Tab. 2: Eigenschaften von DWDS-KORPUS

Funktionalität	Linguistische Forschung
Sprachenauswahl	Monolingual: Deutsch
Medium	Geschriebene Sprache und gesprochene Sprache in einigen Korpora (u.a. <i>Wendekorpus</i>). Die Medialität ist in einigen Korpora unterschiedlich, da man dort auch konzeptuell mündliche Korpora finden kann (u.a. <i>Blogs</i>).
Größe	27 Milliarden Tokens (Stand 02.02.2021).
Annotation	Annotation (automatisch).
Persistenz	Auch als Monitorkorpus nutzbar.

Untersuchungsgegenstand	Referenzkorpus. Einige Korpora sind Spezialkorpora.
Verfügbarkeit	Frei ohne Registrierung. Einige Korpora mit Registrierung

Das *Digitale Wörterbuch der deutschen Sprache* (DWDS)¹² ist ein ‚Wortinformationssystem‘, das öffentlich zugänglich ist, aber zur Benutzung spezifischer Korpora und größerer Datenmengen ist eine Registrierung notwendig.

Es wird als ausgewogenes Korpus der deutschen Sprache des 20. Jahrhunderts definiert und es enthält unterschiedliche Typen von Korpora: *Referenzkorpora*, die zeitlich und hinsichtlich der Textsortenverteilung ausgewogen sind, wie das *DWDS-Kernkorpus des 20. Jahrhunderts*, das *DWDS-Kernkorpus des 21. Jahrhunderts*, das *DTA-Kernkorpus Metakorpora* (d.h. mehrere Korpora, die zur einfachen Recherchen aggregiert wurden, wie *Historische Korpora* etc.); *Zeitungskorpora*, die retrodigitalisierte und digitale Zeitungstexte aus deutschen Tageszeitungen und Wochenzeitungen enthalten, wie *Berliner Zeitung*, *Der Tagesspiegel*, *Die Zeit* (ab 1946), *Neues Deutschland* (1946–1999); *Webkorpora*, d.h. Korpora, die aus internetbasierten Quellen entstanden sind, wie das *Webkorpus*, das *Jurakorpus*, das *Medizinkorpus*, das *Corona-Korpus* etc.; *Spezialkorpora*, wie *Dinglers Polytechnisches Journal*, *Filmuntertitel*, *Gesprochene Sprache*, *Text+Berg*, *Berliner Wendekorpus*, *Politische Reden* (1982–2020) etc.

Alle Korpustexte verfügen über Metadaten zu Textsorte, Autor, Titel, Verlag und Erscheinungsdatum und sind linguistisch annotiert.

In der DWDS-Abfrageplattform sind unterschiedliche Werkzeuge enthalten, die auch für den DaF-Bereich von großer Bedeutung sind:

- die linguistische Suchmaschine *DDC*, die eine Recherche nach Wortform, Lemma und Wortart ermöglicht und deren Ergebnisse nach unterschiedlichen Kriterien und Formaten sortiert werden können:

¹² Vgl. <https://www.dwds.de>

Startseite / Wörterbuch / Corona-Pandemie – Schreibung, Definition, Bedeutung, Synonyme, Beispiele

Corona-Pandemie

Corona-Pandemie, die

Alternative Schreibung **Coronapandemie**
Grammatik Substantiv (Femininum) · Genitiv Singular: **Corona-Pandemie** · Nominativ Plural: **Corona-Pandemien** · Verwendung im Plural ungebräuchlich
Aussprache [koˈroːnapandɛ.miː]
Worttrennung Co-ro-na-Pan-de-mie • Co-ro-na-pan-de-mie (computergeneriert)
Wortzerlegung /Corona /Pandemie
Rechtschreibregeln § 45 (1)

Dieses Stichwort finden Sie im DWDS-Themenglossar zur COVID-19-Pandemie.

Bedeutung

durch das **Coronavirus (2)** im Jahr 2020 ausgelöste weltweite schwere **Epidemie** ZDL-Vollartikel

BEISPIELE:
Der Spielbetrieb ruht, Sich-Treffen ist, von den bekannten Ausnahmen abgesehen, verboten und die Vorstandsarbeit damit von einer ganz anderen Qualität als in der Zeit, als noch niemand wusste, was eine **Corona-Pandemie** bedeutet. [Allgemeine Zeitung, 19.12.2020]
Sie [eine Journalistin] wollte erklären, woher Angst kommt, nicht nur im Krieg, sondern auch in relativer Sicherheit, im westlichen Alltag, in Krisen wie der **Corona-Pandemie**. [Süddeutsche Zeitung, 19.12.2020]
Wegen der **Coronapandemie** verzichten Kirchgemeinden in Radebeul auf den traditionellen Gottesdienst am Heiligen Abend. [Dresdner Neueste Nachrichten, 18.12.2020]
... **3 weitere Belege**

letzte Änderung: 20.01.2021

Thesaurus

www.openthesaurus.de (01/2021)

Synonymgruppe
COVID-19-Pandemie · Corona-Pandemie · Coronavirus-Krise · Coronavirus-Pandemie •
/Coronakrise ugs., Hauptform

Zitationshilfe

„Corona-Pandemie“, bereitgestellt durch das Digitale Wörterbuch der deutschen Sprache, <<https://www.dwds.de/wb/Corona-Pandemie>>, abgerufen am 12.02.2021.
Weitere Informationen ...

alphabetisch vorangehend	alphabetisch nachfolgend
Corona-Krise	Corona-Party
Corona-Kabinett	Corona-Radweg
Corona-Impfstoff	Corona-Sünder
Corona-Ferien	Corona-Sünderin
Corona-Demo	Corona-Tracing-App

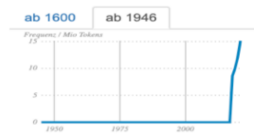
Worthäufigkeit

selten häufig

Wortverlaufskurve

ab 1600 ab 1946

Prozente / Mio. Wörter



Weitere Wörterbücher

- Deutsches Wörterbuch (DWB) (0)
- Deutsches Wörterbuch, Neubearbeitung (DWB) (0)
- Wörterbuch der deutschen Gegenwartssprache (WDG) (0)

Belege in Korpora

Referenzkorpora

- DWDS-Kernkorpus (1900–1999)
- DWDS-Kernkorpus 21 (2000–2010)
- DTA-Kernkorpus (1598–1913)

Metakorpora

- DTA-Kern+Erweit. (1465–1969)
- Historische Korpora (1465–1969)
- Referenz- und Zeitungskorpora (frei)

Zeitungskorpora

- Berliner Zeitung (1994–2005)
- Der Tagesspiegel (ab 1996)
- Die ZEIT (1946–2018)

Webkorpora

- Blogs

Spezialkorpora

- DTA-Erweiterungen (1465–1969)
- Archiv der Gegenwart (1931–2000)
- Polytechnisches Journal
- Filmuntertitel
- Gesprochene Sprache
- DDR
- Politische Reden (1982–2020)

Abb. 8: Ergebnisse der Suche nach *Corona-Pandemie*

Wie man Abbildung 8 entnehmen kann, kann eine Vielfalt an Informationen über das gesuchte Wort *Corona-Pandemie* entnommen werden, die für den DaF-Unterricht von großer Bedeutung sind: Häufigkeit des Wortes, Alternativschreibung (*Corona-Pandemie/Coronapandemie*), grammatische Informationen, Aussprache, Worttrennung, Wortzerlegung, Rechtschreibung, Informationen zur Bedeutung des Wortes mit entsprechenden Belegen, Synonyme, Komposita, Hinweis auf die Wortverlaufskurve, Belege in Wörterbüchern und Korpora;

- *Wortprofil*, zur Extraktion von Kollokationen in synchroner Perspektive. Kookkurrenzpartner können nach unterschiedlichen statistischen Maßen extrahiert werden, so dass ein kompakter Überblick über statistisch signifikante Wortverbindungen gegeben werden kann:

Suche im DWDS-Wortprofil

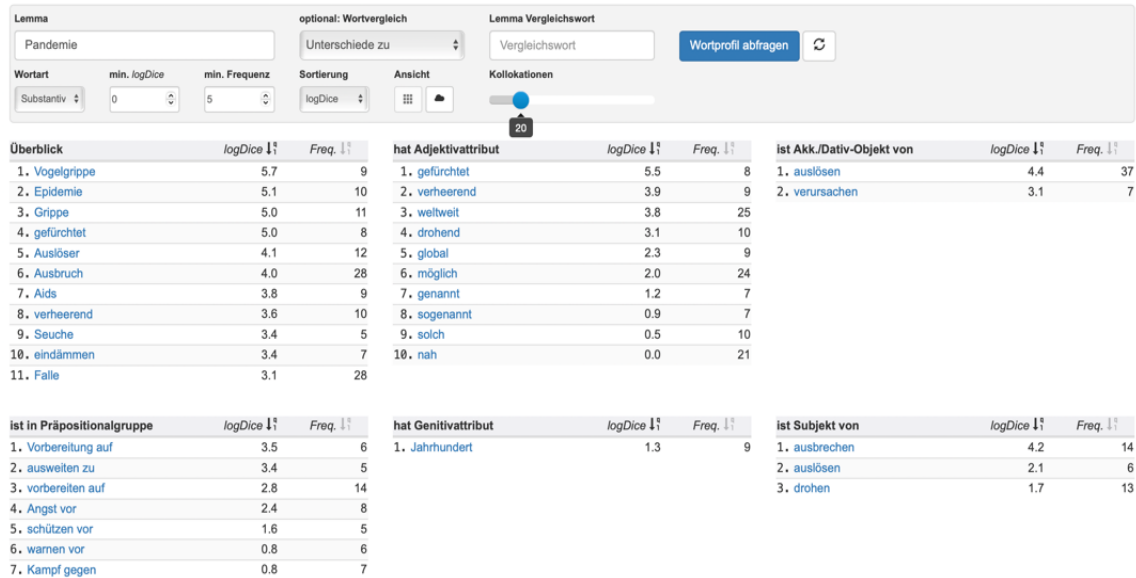


Abb. 9: Ergebnisse des Wortprofils von *Pandemie*

Das Substantiv *Pandemie* kookkuriiert auf der Basis des Parameters *Logdice* u.a. mit den Adjektivattributen *gefürchtet*, *verheerend*, *weltweit*, *drohend*, *global* oder kommt als Akkusativ-Objekt mit *auslösen* und *verursachen* vor. Es ist auch möglich, ein Wort in Relation zu einem Vergleichswort (z.B. ein partielles Synonym) zu setzen, so dass auf die Bedeutung der jeweiligen Wörter, sowie auf die bevorzugten Kollokationen fokussiert werden kann;

- *DiaCollo* zur Extraktion von Kollokationen in diachroner Perspektive. Mit diesem Tool kann untersucht werden, wie sich die Häufigkeiten im gemeinsamen Vorkommen im Laufe der Zeit verändert haben:



Abb. 10: Ergebnis der Suche nach *Pandemie* in *DiaCollo*

Der Word-Cloud kann zum Beispiel entnommen werden, dass die typischen Wortnachbarn in einem festgelegten Abstand zum Wort *Pandemie* zwischen 2000 und

2010 das Akronym *Who* (*World Health Organization*), die Substantive *Schweinegrippe*, *Impfstoff*, *Virus*, *Bevölkerung*, *Experte*, *Mensch*, *Fall* und die Adjektive *weltweit*, und *neu* sind (Korpus *Die Zeit*). Wie es typisch für Wortwolken ist, entspricht der Größe eines Wortes die Frequenz (je größer das Wort, je höher ist der *Score*, der das gemeinsame Vorkommen der zwei Wörter bestimmt). In dieser Darstellung sind auch Farben wichtig, da Wörter in der gleichen Farbe verdeutlichen, dass die Wörter ähnliche *Scores* haben. Neben der graphischen Darstellung als Wortwolke können auch weitere, wie *Bubble*, *HTML* etc. ausgewählt werden.

Beim Klicken auf eins der Wörter in der Wortwolke, wie z.B. auf *Schweinegrippe* öffnen sich die jeweiligen KWICs, so dass die Beispiele entnommen werden können:



The screenshot shows a search interface with the following content:

D*/zeit_www Search
 Hits 1 - 6 of 6

Navigation: ~HTML | ~Hist | Home | Query Lizard | Previous | Next | Help

Search Query: NEAR((\$!=@'Pandemie') =1, (\$!=@'Schweinegrippe' WITH \$p=@'NN') =2, 8) #SEPARATE #asc_date[2] submit +

Results:

- 1: [zeit:part1/2009/05/12/amerikagrippe_p...] ... der **Schweinegrippe**_[2] geht ein erhebliches Potenzial für eine **Pandemie**_[1] aus.
- 2: [zeit:part1/2009/05/21/schweinegrippe...] ... Kapazitäten, falls zu der üblichen Grippewellen eine **Pandemie**_[1] der sogenannten **Schweinegrippe**_[2] hinzukäme.
- 3: [zeit:part1/2009/06/11/amerikagrippe_w...] Die sogenannte **Schweinegrippe**_[2] ist eine **Pandemie**_[1] .
- 4: [zeit:part1/2009/06/18/Kommentar_Gripp...] ... grassierende **Schweinegrippe**_[2] zur weltweiten Seuche erklären – zur **Pandemie**_[1] .
- 5: [zeit:part1/2009/09/11/impfgegner_schw...] ... gegen die als **Schweinegrippe**_[2] bekannt gewordene und als **Pandemie**_[1] deklarierte Influenza H1N1/09 vor.
- 6: [zeit:part1/2009/11/21/gutenberg_hali...] ... um internationalen Drogenhandel und die Angst vor einer **Pandemie**_[1] namens **Schweinegrippe**_[2].

Navigation: ~HTML | ~Hist | Home | Query Lizard | Previous | Next | Help

Search Query: NEAR((\$!=@'Pandemie') =1, (\$!=@'Schweinegrippe' WITH \$p=@'NN') =2, 8) #SEPARATE #asc_date[2] submit +

Footer: D* OpenSearch API version 0.59 | Imprint | Privacy

Abb.11: Die KWICS von Pandemie mit Schweinegrippe

- *Wortverlaufskurve* zur Erstellung von Frequenzprofilen von Einzelwörtern über bestimmte Zeitabschnitte. Die Häufigkeit eines Wortes kann sich im Laufe der Zeit ändern: Sie kann zunehmen oder abnehmen, ein Wort kann durch ein anderes ersetzt werden, gar nicht mehr benutzt werden oder ganz neu entstehen, wie diese Verlaufskurve des Wortes Corona-Pandemie veranschaulicht:

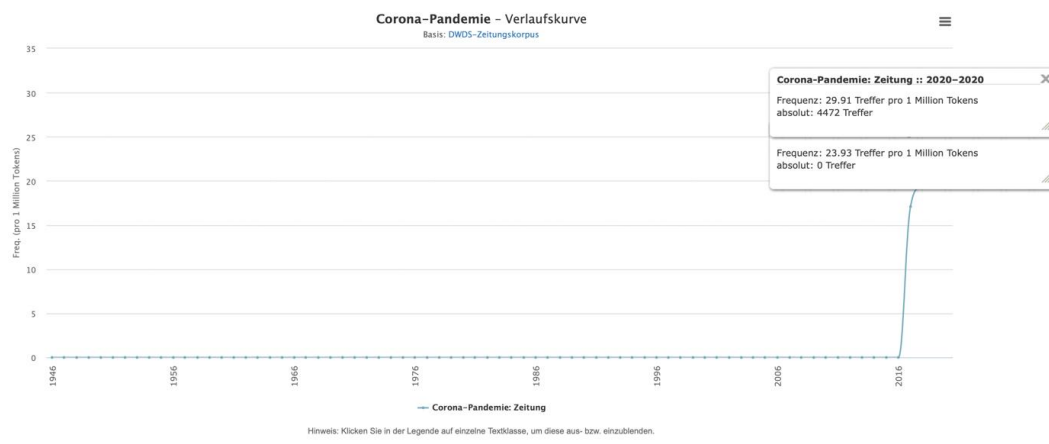


Abb. 12: Wortverlaufskurve des Wortes *Corona-Pandemie*

Benutzeroberfläche und Suchfunktionen der Suchmaschine (nach Wortform, Lemma und Wortart) sind leicht handhabbar, und die Ergebnisse können nach unterschiedlichen Kriterien und Formaten sortiert werden (vgl. Steyer 2008: 190). Die möglichen Anwendungen des DWDS Wortinformationssystem im DaF-Unterricht wurden schon in mehreren Aufsätzen veranschaulicht (vgl. u.a. Wallner 2013; Flinz/Katelhön 2019; Krekeler in dieser Themenausgabe; Malloggi 2021).

3.3 Das Korpus German Web 2013

Tab. 3: Eigenschaften von German Web 2013

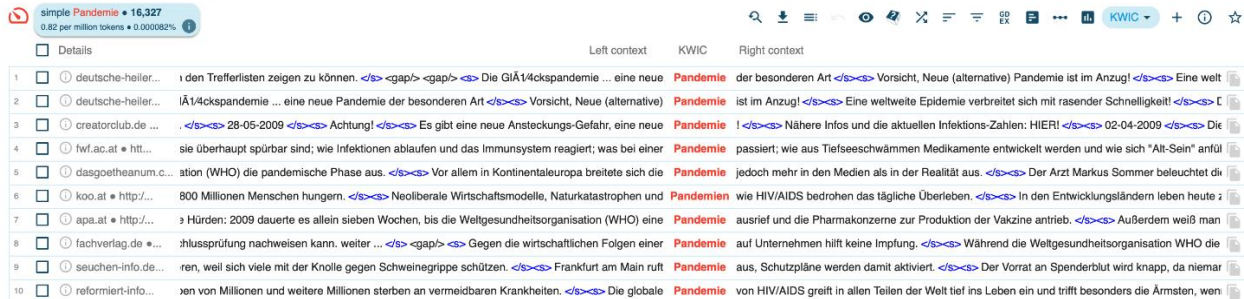
Funktionalität	Linguistische Forschung
Sprachenauswahl	Monolingual: Deutsch
Medium	Geschriebene Sprache
Größe	20 Milliarden Tokens (Stand 02.02.2021)
Annotation	Annotation (automatisch)
Persistenz	Statisches Korpus
Untersuchungsgegenstand	Referenzkorpus
Verfügbarkeit	Frei bis 2022 mit Registrierung

Das Korpus *German Web 13 (deTenTen13)* ist ein automatisch generiertes Internetkorpus aus dem Jahr 2013 mit 16.5 Billionen Wörtern (86,85% deutsche Texte, 6,54% österreichische Texte, 6,24% schweizerische Texte; 0,38% Texte aus anderen deutschsprachigen Ländern). Das Korpus¹³ ist automatisch lemmatisiert und morphosyntaktisch annotiert (RFT-Tagger) und befindet sich in der Sketch Engine Webanwendung, das dank eines europäischen Projektes bis 2022 nach Registrierung für Universitäten und Forschungseinrichtungen frei zugänglich ist.

Das Korpus deTenTen13 kann mit folgenden Funktionalitäten analysiert werden:

- *Concordance* zur Untersuchung von Textwörtern oder Lemmata (auch mit Benutzung der *Corpus Query Language*) in ihrem Kontext:

¹³ In der Sketch Engine-Anwendung können auch eigene Korpora erstellt werden, die auch interessante Auswirkungen auf DaF haben können (auch wenn Fachsprachen das Thema sind) (vgl. dazu Flinz 2020).



Details	Left context	KWIC	Right context
1	deutsche-heiler... den Trefferlisten zeigen zu können.	Pandemie	der besonderen Art
2	deutsche-heiler... lä14ckspandemie ... eine neue	Pandemie	ist im Anzug!
3	creatorclub.de ... 28-05-2009	Pandemie	Nähere Infos und die aktuellen Infektions-Zahlen: HIER!
4	fwf.ac.at ... sie überhaupt spürbar sind; wie Infektionen ablaufen und das Immunsystem reagiert; was bei einer	Pandemie	passiert; wie aus Tiefseeschwämmen Medikamente entwickelt werden und wie sich "Alt-Sein" anfü
5	dasgoethearum.c... ation (WHO) die pandemische Phase aus.	Pandemie	jedoch mehr in den Medien als in der Realität aus.
6	koo.at ... 800 Millionen Menschen hungern.	Pandemie	wie HIV/AIDS bedrohen das tägliche Überleben.
7	apa.at ... Hürden: 2009 dauerte es allein sieben Wochen, bis die Weltgesundheitsorganisation (WHO) eine	Pandemie	ausrief und die Pharmakonzerne zur Produktion der Vakzine antrieb.
8	fachverlag.de ... hlussprüfung nachweisen kann. weiter ...	Pandemie	auf Unternehmen hilft keine Impfung.
9	seuchen-info.de ... ren, weil sich viele mit der Knolle gegen Schweinegrippe schützen.	Pandemie	aus, Schutzpläne werden damit aktiviert.
10	reformiert-info... en von Millionen und weitere Millionen sterben an vermeidbaren Krankheiten.	Pandemie	von HIV/AIDS greift in allen Teilen der Welt tief ins Leben ein und trifft besonders die Ärmsten, wen

Abb. 13: KWICs des Suchwortes *Pandemie* mit *Concordance*

In Abbildung 13 wird das Suchwort *Pandemie* in seinem linken und rechten Kontext veranschaulicht (normalerweise 5 Wörter links und 5 Wörter rechts). Der Kontext kann auch zusätzlich nach links oder rechts gefiltert werden, so dass zum Beispiel alle Wörter davor oder danach fokussiert werden können:



Details	Left context	KWIC	Right context
2669	welt-diabetes-t... 2006 eine Resolution verabschiedet und die Krankheit als Volkskrankheit und weltweit bedrohliche	Pandemie	eingestuft.
2670	enutrio.de ... Diese Worte klingen fast wie eines der zehn Gebote und sind – betrachtet man die bedrohliche	Pandemie	des Übergewichts – wohl auch genauso schwer einzuhalten.
2671	heute-gesund-le... die durch den A-H1N1-Virus hervorgerufene neue Grippe.	Pandemie	ansahen, hielten andere für einen aufgebauchten Marketing-Gag der Pharmaindustrie.
2672	gutepillen-schi... e höchste Stufe ausgerufen werden konnte – ohne dass tatsächlich Hinweise auf eine bedrohliche	Pandemie	bestandent sie das erlaubten.
2673	promed-ev.de ... id auszuweisen.	Pandemie	vorzubereiten?
2674	heute-gesund-le... die durch den A-H1N1-Virus hervorgerufene neue Grippe.	Pandemie	ansahen, hielten andere für einen aufgebauchten Marketing-Gag der Pharmaindustrie.
2675	apotheker.or.at... Zigarettenkonsum stark steigend.	Pandemie	. Man kann annehmen, dass ca. 30% der Krebsstodesfälle, 30-40% der Todesfälle durch Herz-Kreis
2676	digital-nippon... ikaverbrauch deutlich herabzusetzen, ansonsten züchten wir uns die nächste, wirklich bedrohliche	Pandemie	selbst.
2677	konsolen-oldies... ap/ - Zwei Krankheitswellen zogen über den Erdball, eine davon wurde sogar als bedrohliche	Pandemie	eingestuft (Vogel- und Schweinegrippe) - Die deutsche Bundesregierung rief Neuwahlen a
2678	enutrio.de ... Diese Worte klingen fast wie eines der zehn Gebote und sind – betrachtet man die bedrohliche	Pandemie	des Übergewichts – wohl auch genauso schwer einzuhalten.
2679	mitteldeutsche... kürzlich in BMC Medicine (2009; 7: 81) darlegten.	Pandemie	eine Quarantäne vor die auf die Passagiere in der Economy Class begrenzt werden könnte.
2680	medicummundi.ch... it Swasiland mit rund einer Million Einwohnerinnen und Einwohnern hat sich Aids zur bedrohlichen	Pandemie	ausgeweitet: Insgesamt ist jeder dritte Erwachsene mit dem HI-Virus infiziert.
2681	kirchenburgen.d... beginnt, darf noch etwas gerätselt werden wie dieser tobringende Virus zur weltweit bedrohlichen	Pandemie	wird.
2682	swissinfo.ch ... ist es wichtig, sich mit der Krankheit auseinanderzusetzen, um einer allfällig wirklich bedrohlichen	Pandemie	entgegen zu wirken.
2683	strohballenbau... it sich offensichtlich um eine ziemlich harmlose Grippevariante.	Pandemie	hochzustilisieren, war nur möglich, weil die WHO rechtzeitig die Bedingungen dafür änderte.
2684	musiker-board.d... as Virus mutiert und zwar mit dem H5N1.	Pandemie	entwickeln (bzw. die Pandemie haben wir ja jetzt schon, aber der jetzige Virus ist nicht bedrohliche
2685	socialnet.de ... mangel, Bodenerosion, Ausbeutung der Rohstoffe, bis hin zur Armut in der Welt, den bedrohlichen	Pandemie	und dem gefährlichen Ungleichgewicht der Ressourcenverteilung - ihre Ursachen in der Umwelt he
2686	infekt.ch ... sssartiger Erfolg zu werten.	Pandemie	des laufenden Jahrhunderts einen gesichert wirksamen Impfstoff zur Verfügung haben, noch bevoi
2687	nlnv.de ... k oder Medien mit verschiedensten Horrorszenerarien beschäftigt.	Pandemie	bislang ausgeblieben, einige medizinische Handlungsoptionen sind in der Zwischenzeit jedoch sch
2688	heute-gesund-le... eraus hohen Anstieg der Infektionsrate von Grippe war auch die Schweinegrippe.	Pandemie	ließ sowohl Patienten als auch Ärzte besonders vorsichtig werden: "Traten die klassischen Grippe
2689	flv-bb.de ... Krisenmanagement	Pandemie	und die möglichen Auswirkungen terroristischer Anschläge in Deutschland zeigen die Notwendigke

Abb. 14: Orientierung der KWICs des Suchwortes *Pandemie* nach dem linken Kontext (*Concordance*)

Wenn die KWICs nach dem linken Kontext orientiert sind, können zum Beispiel die häufigsten Adjektive, die in der attributiven Funktion mit *Pandemie* benutzt werden, identifiziert werden. Aus der Abbildung 14 können zum Beispiel die Adjektive *bedrohlich* und *befürchtet* entnommen werden.

- *Word List* zur Erstellung von Listen auf der Basis der absoluten Häufigkeit (Wörter, Lemmata, Substantive, Verben etc.)

WORDLIST German Web 2013 (deTenTen13) Se non riesci a vedere l'email clicca qui Segnala Spam 2 pro

word (13,781,354 items | 19,708,225,694 total frequency)

Word	Frequency ?	Word	Frequency ?	Word	Frequency ?	Word	Frequency ?
1 ,	909,061,297 ...	11 zu	173,223,679 ...	21 auch	122,708,712 ...	31 -	88,924,960 ...
2 .	864,497,563 ...	12 mit	168,140,926 ...	22 :	120,189,897 ...	32 an	77,936,704 ...
3 die	527,083,497 ...	13 ist	162,143,640 ...	23 sich	118,019,839 ...	33 als	77,443,511 ...
4 und	488,825,209 ...	14 von	154,641,825 ...	24 ich	116,629,226 ...	34 werden	71,710,411 ...
5 der	477,363,116 ...	15 für	132,401,619 ...	25 sie	115,401,847 ...	35 bei	71,360,429 ...
6 in	267,462,955 ...	16 auf	128,535,376 ...	26 eine	110,052,212 ...	36 sind	64,130,080 ...
7 "	219,101,466 ...	17 nicht	125,766,640 ...	27)	99,356,616 ...	37 dass	63,666,617 ...
8 <gap/>	204,448,531 ...	18 es	125,616,825 ...	28 dem	96,322,300 ...	38 oder	63,638,390 ...
9 das	201,695,204 ...	19 ein	124,531,445 ...	29 des	95,347,159 ...	39 so	63,602,553 ...
10 den	179,265,126 ...	20 im	122,772,780 ...	30 (94,007,761 ...	40 aus	63,541,047 ...

You are only allowed to access 1,000 items. [Get more](#) Rows per page: 50 1-50 of

Abb. 15: Wordlist (Parameter Word) in German Web 2013

Wie es typisch für Wortlisten in allgemeinsprachlichen Korpora ist, kommen an erster Stelle Interpunktionszeichen¹⁴, Artikel, Konjunktionen, Auxiliärverben etc. Autosemantika, wie Substantive, Adjektive etc. kommen erst später vor;

- *Keywords* zur Extrahierung von einzelnen *Keywords* und *Terms* (*multiple Keywords*) auf der Basis eines Referenzkorpus¹⁵;
- *N-Grams* zur Extrahierung von N-Grammen (Bigramme, Trigramme etc.), die die Aufmerksamkeit auf Phänomene der semantischen und pragmatischen Ebene lenken können;
- *Word Sketches* zur automatischen Zusammenfassung des grammatischen und kollokationellen Verhaltens eines Wortes (vgl. Abbildung 16):

WORD SKETCH German Web 2013 (deTenTen13)

Pandemie as noun 16,064x ...

Constructions	modifiers of "Pandemie"	verbs with "Pandemie" as subject	genitive objects of "Pandemie"
"%w" being regular: 16,064 ...	befürchtet: 31 ... die befürchtete Pandemie	ausbrechen: 39 ... eine Pandemie ausbricht	Aids: 14 ... der HIV / Aids Pandemie
"%w" in nominative: 4,988 ...	ausbreitend: 15 ... ausbreitenden Pandemie	bevorstehen: 17 ... Pandemie bevorsteht	H1N1: 13 ... H1N1 Pandemie
"%w" in accusative: 3,920 ...	ausgerufen: 11 ... ausgerufenen Pandemie	wüten: 10 ...	Schweinegrippe: 41 ... Pandemie der Schweinegrippe
"%w" in dative: 3,821 ...	grassierend: 12 ... grassierende Pandemie	breiten: 25 ... Pandemie breitet sich	Grippe: 26 ... Pandemie der Spanischen Grippe
"%w" in genitive: 3,335 ...	weltumspannend: 13 ... weltumspannenden Pandemie	ausbleiben: 6 ...	Influenza: 15 ... Influenza Pandemie
	voreilig: 16 ... die Schweinegrippe voreilig zur Pandemie erklärt zu haben	bedrohen: 6 ...	Vogelgrippe: 17 ... Pandemie der Vogelgrippe
	drohend: 155 ... einer drohenden Pandemie	verbreiten: 13 ...	Ausbruch: 18 ... Ausbruchs einer Pandemie
	selektiv: 40 ... der selektiven Pandemie	drohen: 40 ... Pandemie droht	Epidemie: 6 ...
	ausgebrochen: 7 ...	brechen: 29 ...	Menschheitsgeschichte: 9 ...
	todbringend: 6 ...	aufreten: 8 ...	WHO: 8 ...
	gefürchtet: 22 ... die gefürchtete Pandemie	verlaufen: 22 ...	Impfstoff: 6 ...
	tödlich: 93 ... eine tödliche Pandemie		Ausmaß: 13 ...

Abb. 16: Ausschnitt des Word Sketch von *Pandemie*

¹⁴ Interpunktionszeichen kommen in Häufigkeitslisten vor, wenn sie als Token gezählt werden. Zu diesem Thema vgl. u.a. Stefanowitsch 2020; Perkuhn/Keibel/Kupietz 2012: 30–32.

¹⁵ Zur Identifizierung wird *Simple Maths* verwendet.

Aus dem *Word Sketch* eines Wortes können seine typischen Wortverbindungen auf der Basis von grammatischen Relationen identifiziert werden. Wenn wir zum Beispiel wissen möchten, welche die typischsten adjektivischen Attribute sind, die sich mit *Pandemie* verbinden, können wir sie aus der Spalte „modifiers of *Pandemie*“ entnehmen: *befürchtete Pandemie*, *ausbreitende Pandemie*, *ausgerufene Pandemie* etc. Beim Klicken auf das Adjektiv öffnen sich auch die KWICs (Abbildung 17):



Abb. 17: KWICs von *Pandemie* mit *befürchtet*

- *Word Sketch Difference* zur Visualisierung des kollokationellen Profils von zwei Wörtern auf der Basis grammatischer Relationen. Diese Funktionalität ist besonders bei Wörtern in semantischer Relation (z.B. Synonyme) interessant:



The image shows the 'WORD SKETCH DIFFERENCE' interface comparing 'Pandemie' (16,064x) and 'Seuche' (47,513x). The interface displays five columns of grammatical relations, with cells colored green for 'Pandemie' and red for 'Seuche'. Grey cells indicate relations shared by both words.

modifiers of "Pandemie/Seuche"	nouns with "Pandemie/Seuche" as genitive object	genitive objects of "Pandemie/Seuche"	dative objects of "Pandemie/Seuche"	nouns with "Pandemie/Seuche" as dative object
befürchtet 31 0 ...	Ausrufung 74 0 ...	Aids 14 0 ...	H1N1 21 0 ...	Schweinegrippe 37 0 ...
ausgerufen 11 0 ...	H1N1 28 0 ...	H1N1 13 0 ...	Schweinegrippe 24 0 ...	Ausrufung 31 0 ...
voreilig 16 0 ...	Ausrufen 18 0 ...	Schweinegrippe 41 0 ...	HIV 8 0 ...	Ausrufen 11 0 ...
drohend 155 40 ...	HIV 16 0 ...	Grippe 26 0 ...	Vogelgrippe 10 0 ...	Epidemie 10 0 ...
weltumspannend 13 13 ...	Impfstoff 26 0 ...	Influenza 15 0 ...	Epidemie 9 0 ...	Grippe 11 0 ...
ausbreitend 15 73 ...	Ernstfall 11 0 ...	Vogelgrippe 17 0 ...	Impfstoff 11 0 ...	Vogelgrippe 12 7 ...
todbringend 6 76 ...	Eindämmung 46 104 ...	Ausbruch 18 56 ...	Aids 8 29 ...	Eindämmung 42 70 ...
grassierend 12 170 ...	Ausbruch 230 725 ...	Nurgles 0 6 ...	Einhalt 0 37 ...	Ausbruch 129 365 ...
hochansteckend 0 56 ...	Ausbreitung 63 467 ...	Nurgle 0 8 ...	Pocke 0 9 ...	Ausbreitung 16 66 ...
ansteckend 0 179 ...	Verschleppung 0 34 ...	Ausbreitung 0 48 ...	Typhus 0 13 ...	Einschleppung 0 14 ...
anzeigepflichtig 0 83 ...	Einschleppung 0 39 ...	Pestilenz 0 10 ...	Cholera 0 31 ...	Pest 0 29 ...
eingeschleppt 0 107 ...	Ausbreiten 0 45 ...	Untod 0 20 ...	Befallene 0 6 ...	Cholera 0 28 ...

Abb. 18: Word Sketch von *Pandemie* und *Seuche*

Wenn wir das Tool für den Vergleich von *Pandemie* und *Seuche* benutzen (Abbildung 18), werden die typischen Kollokationen zu jedem Wort farbig hervorgehoben: grün für *Pandemie* und rot für *Seuche*. Je stärker die Farbe ist, je typischer ist die Verbindung (vgl. *befürchtete Pandemie* und *ansteckende Seuche*); die grauen Kollokatoren können hingegen für beide Wörter benutzt werden: *weltumspannend* kann sowohl mit *Pandemie* als auch mit *Seuche* verwendet werden;

- *Thesaurus* zur Visualisierung von Synonymen und ähnlichen Wörtern auf der Basis des Kollokationsprofils (Abbildung 19):

	Word	Frequency ?
1	Epidemie	29,398 ...
2	Schweinegrippe	39,364 ...
3	Seuche	47,513 ...
4	Vogelgrippe	24,816 ...
5	Influenza	20,628 ...
6	Umweltkatastrophe	10,239 ...
7	Cholera	21,094 ...
8	Hungersnot	33,274 ...
9	H1N1	9,262 ...
10	Tierseuche	7,411 ...

Abb. 19: Thesaurus zu *Pandemie* (tabellarische Darstellung)

Dem Thesaurus zu *Pandemie* kann entnommen werden, dass folgende Wörter ein ähnliches Kollokationsprofil haben: *Epidemie*, *Schweinegrippe*, *Seuche*, *Vogelgrippe*, *Influenza* etc. Der *Thesaurus* kann auch mit folgender Visualisierung dargestellt werden (die Zahl der Treffer wurde auf zehn reduziert):

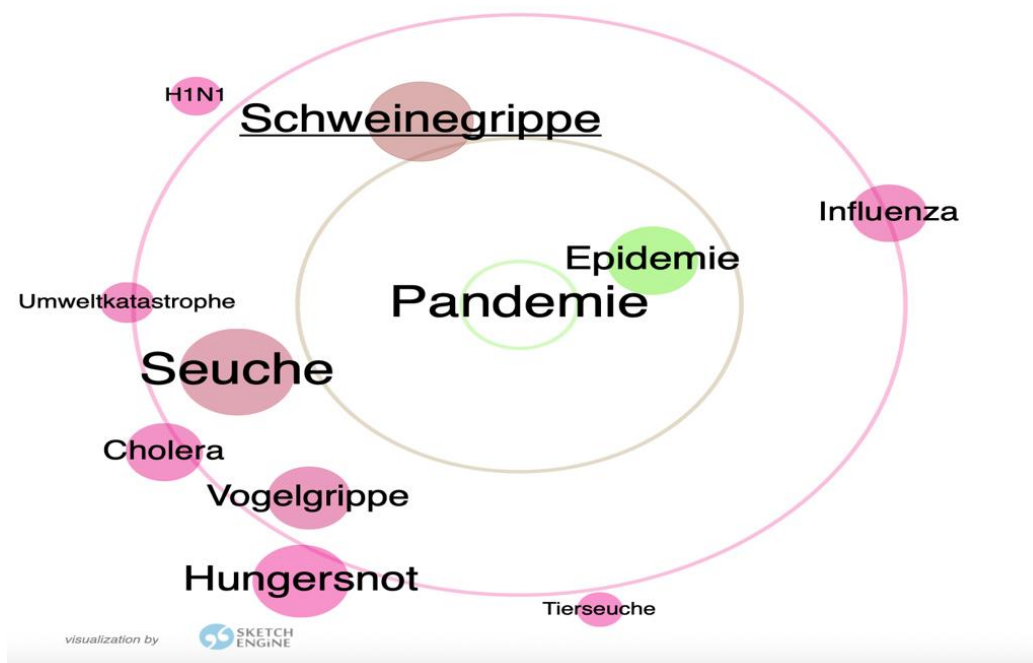


Abb. 20: Thesaurus zu *Pandemie* (Visualisierung)

Anwendungen der Sketch Engine Umgebung im DaF-Unterricht sind noch in ihren Anfängen (vgl. u.a. Flinz 2020).

3.4 Das Korpus FOLK

Tab. 4: Eigenschaften von FOLK

Funktionalität	Linguistische Forschung
Sprachenauswahl	Monolingual: Deutsch
Medium	Gesprochene Sprache
Größe	2.719.948 Textwörter (Stand 02.02.2021)
Annotation	Annotation (automatisch)
Persistenz	Auch als Monitorkorpus nutzbar
Untersuchungsgegenstand	Referenzkorpus
Verfügbarkeit	Frei mit Registrierung

Das Korpus *FOLK* (Forschungs- u. Lehrkorpus für gesprochenes Deutsch) ist ein Korpus der gesprochenen Sprache, das sich an Forschende, Lehrende und Studierende aus den Bereichen der Gesprächsforschung, der Korpuslinguistik sowie damit verwandten Ansätzen richtet¹⁶. Es besteht aus 332 Gesprächsereignissen mit 1131 dokumentierten SprecherInnen in unterschiedlichen Kontexten: private (wie Tischgespräche, Gespräche bei Spielinteraktionen), institutionelle (wie Schulunterricht, berufliche Kommunikation) und öffentliche (wie Podiumsdiskussionen, Schlichtungsgespräche). Von den 332 Gesprächsereignissen sind für 105 Gesprächsereignisse auch Videoaufnahmen vorhanden. Die Gespräche weisen bezüglich des Parameters *Interaktionstyp* eine große Streuung auf und beim Korpusausbau sind Stratifikationsparameter wie regionale Herkunft, Alter oder Bildungsstand der SprecherInnen berücksichtigt worden, so dass auch die Zusammenstellung von ausgewogenen Teilkorpora möglich ist. Grundlage für die Gesprächsbeschreibung, Metadokumentation, Suchfunktionalitäten und Ausbauplanung ist ein detailliertes Stratifikationskonzept (vgl. u.a. Deppermann/Hartung 2011; Kaiser 2018: 516).

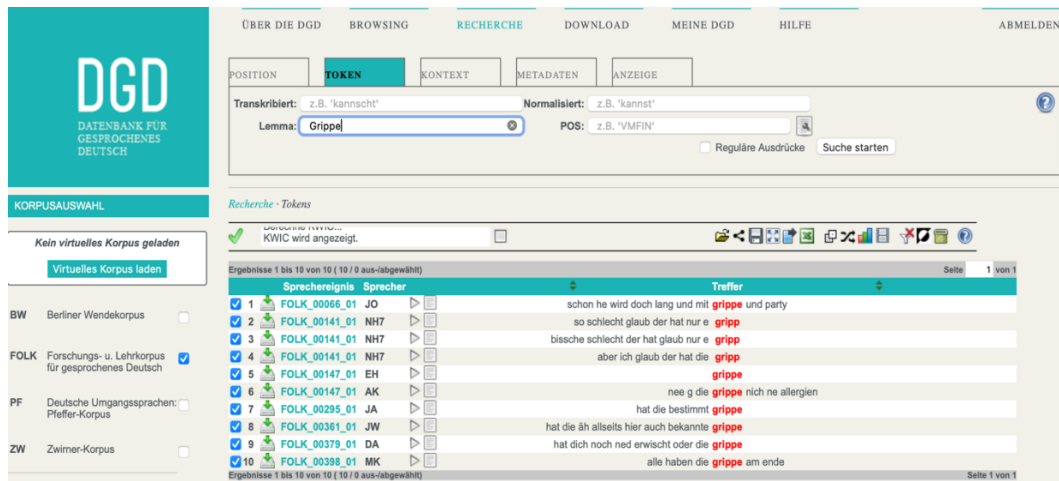
Die Aufnahmen sind nach den Konventionen für computergestützte GAT-Minimaltranskripte in literarischer Umschrift mit dem Editor FOLKER transkribiert worden¹⁷; sie sind in Segmenten von maximal 5 Sekunden mit den Aufnahmen aligniert. Neben der literarischen Transkription sind drei weitere Annotationsebenen hinzugefügt worden: die orthographische Normalisierung, die Lemmatisierung und

¹⁶ Vgl. https://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.welcome

¹⁷ Die Transkripte haben einen Umfang von 2.719.948 laufenden Textwörtern (Version 2.14, Stand April 2020).

das *Part-Of-Speech-Tagging* gemäß einer für Gesprächsdaten erweiterten Version des *Stuttgart-Tübingen-Tagset*.

Das Korpus kann mit den Funktionalitäten der *Datenbank für Gesprochenes Deutsch* (DGD) durchsucht werden, die nach kostenfreier Registrierung zugänglich sind. Recherchen können sowohl in den Transkripten als auch in den Metadaten durchgeführt werden; weiteres Zusatzmaterial (Informationen über Settings und Verläufe der aufgezeichneten Interaktionen, alphabetisch und nach Frequenzen geordnete Wort- und Lemmalisten, Transkriptionskonventionen etc.) steht auch frei zur Verfügung. Mit der DGD kann explorativ nach Formen gesucht werden, und es können auch gezielt formbestimmte Suchanfragen erstellt werden (vgl. Deppermann/Schmidt 2014). Volltextsuchen sowie struktursensitive Suchrecherchen sind möglich und verschiedene Filter bieten die Möglichkeit die Suchergebnisse, die in KWICs dargestellt werden, zu verfeinern.



The screenshot shows the DGD search interface. The search bar contains the lemma 'Grippe'. The search results are displayed in a table with columns for 'Sprecher/Signale', 'Sprecher', and 'Treffer'. The results show various instances of the word 'grippe' in different contexts, such as 'schon he wird doch lang und mit grippe und party' and 'alle haben die grippe am ende'.

Ergebnisse 1 bis 10 von 10 (10 / 0 aus-abgewählt)	Sprecher/Signale	Sprecher	Treffer
1	FOLK_00066_01	JO	schon he wird doch lang und mit grippe und party
2	FOLK_00141_01	NH7	so schlecht glaub der hat nur e grippe
3	FOLK_00141_01	NH7	bissche schlecht der hat glaub nur e grippe
4	FOLK_00141_01	NH7	aber ich glaub der hat die grippe
5	FOLK_00147_01	EH	grippe
6	FOLK_00147_01	AK	nee g die grippe rich ne allergien
7	FOLK_00295_01	JA	hat die bestimm grippe
8	FOLK_00361_01	JW	hat die äh allseits hier auch bekannte grippe
9	FOLK_00379_01	DA	hat dich noch ned erwischt oder die grippe
10	FOLK_00398_01	MK	alle haben die grippe am ende

Abb. 21: Suche nach *Grippe* (Lemma) in der *DGD* (Korpus FOLK)

Mit der DGD können sowohl quantitative als auch qualitative Arbeitsschritte erstellt werden.

Während viele Korpora in der DGD für Zwecke des Fremdsprachenunterrichts nicht geeignet sind, gilt das nicht für FOLK. Neben der Authentizität der Gespräche, der Ausgewogenheit, dem Vorhandensein von Audio und Videoaufnahmen, ist die Tatsache hervorzuheben, dass komplette Datensätze heruntergeladen werden können und somit die Weiterverarbeitung für die Erstellung von Lerneinheiten begünstigt wird (vgl. Imo/Weidner 2018: 238). FOLK kann „gewinnbringend“ für Lehrzwecke eingesetzt werden (Imo/Weidner 2018: 238) vor allem, wenn Lehrende auch über gesprächsanalytische Vorkenntnisse verfügen. Sowohl produktive als auch rezeptive Fähigkeiten im Bereich der mündlichen Kommunikation können hinsichtlich einer Vielzahl von Aspekten gefördert werden (vgl. Kaiser/Schedl in diesem Themenheft). Didaktische Vorschläge können Sieberg (2016), Dietz

(2021), Kaiser/Schedl in dieser Themenausgabe), Moroni (2021) entnommen werden.

Konkrete Unterstützung für die Benutzung von FOLK im DaF-Unterricht wird mit dem Projekt *ZuMult - Zugänge zu multimodalen Korpora gesprochener Sprache – Vernetzung und zielgruppenspezifische Ausdifferenzierung* (<https://zumult.org>) angestrebt. Mit diesem Tool werden neue Zugangswege entwickelt, die für die Bedürfnisse spezifischer Nutzergruppen optimiert sind. *ZuMult* ist direkt mit der DGD verlinkt und kann die Auswahl der adäquaten Gesprächsereignisse erleichtern (vgl. Fandrych/Meißner/Wallner 2018, 2021; Kaiser/Schedl in diesem Themenheft): Die Vorauswahl der Gesprächsereignisse (*ZuMal*) kann auf der Basis von ausgewählten Parametern (wie u.a. Gesprächstyp, Gesprächsart, Gesprächsthema, Sprachregion, Niveaustufe, Standardnähe, Sprechgeschwindigkeit, Überlappungen etc.) erfolgen (Abbildung 22):

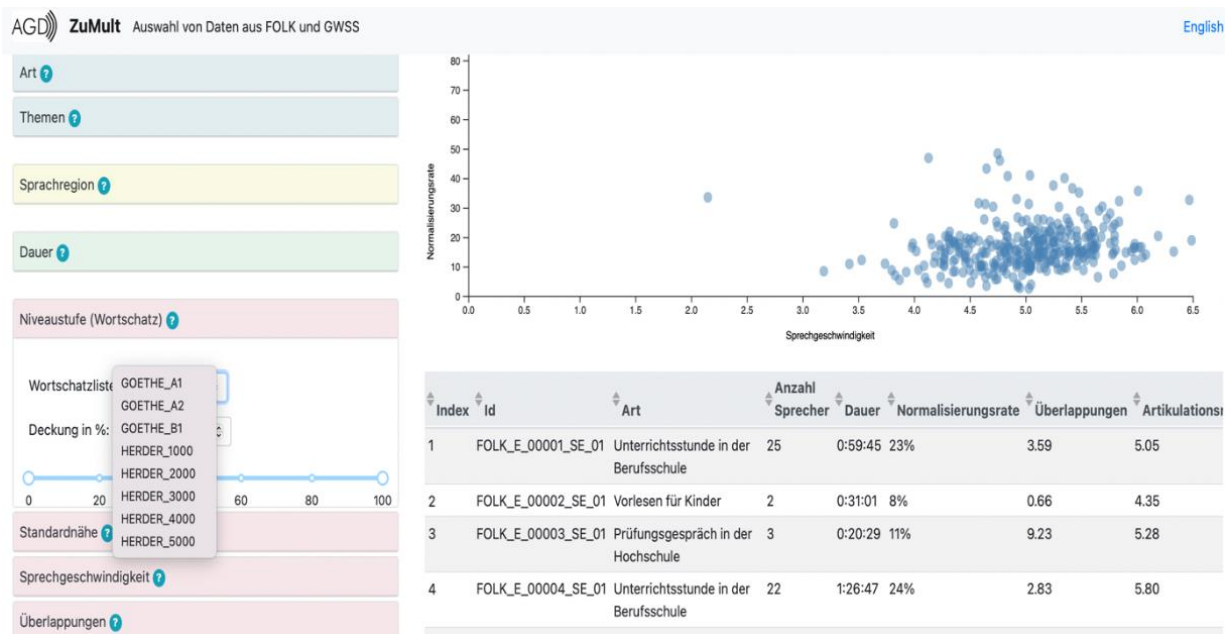


Abb. 22: Auswahl der Gesprächsereignisse nach Niveaustufe

Bei der Suche in FOLK durch *ZuMult* (*Tool Search by Query*) hat man Zugriff auf die Transkripte der DGD. Dank dem in *ZuMult* entwickelten Transkriptbrowser *ZuViel* können relevante Phänomene hervorgehoben werden können.

3.5 Weitere Korpora

Weitere Korpora können der Clarin-Umgebung entnommen werden, wo sie nach Typen kategorisiert sind¹⁸. In jeder Kategorie sind Korpora unterschiedlicher Sprachen enthalten, darunter auch deutsche Korpora. Interessant für DaF können sowohl die Kategorie CMC-Korpora¹⁹ sein, die Chats-, Blogs-, Tweets-, SMS-, und WhatsApp Korpora enthalten, als auch die Kategorie Parallelkorpora²⁰, wenn über Übersetzungsmöglichkeiten reflektiert werden soll. Erwähnenswert in dieser Kategorie sind auch die Parallelkorpora der Europäischen Union (u.a. das *Europarl Parallel Corpus*²¹) sowie das Korpus *InterCorp*²², das als Ziel auch die Unterstützung im Fremdsprachenunterricht angibt. Es enthält manuell alignierte belletristische, populärwissenschaftliche und wissenschaftliche Texte (92.290.000 alignierte Wörter) in mehreren Sprachen (darunter Tschechisch, Deutsch, Englisch, Italienisch etc.). Für einige Sprachen wurden die Korpora morphosyntaktisch annotiert (vgl. <https://intercorp.korpus.cz/?req=page:info>).

Weitere wichtige Korpustypen für DaF sind Vergleichskorpora²³, die noch eine Rarität sind²⁴. Eine Ausnahme bietet das LBC-Korpus (<http://corpora.lessicobenculturali.net/de/>), das Vergleichskorpora in mehreren Sprachen (Deutsch, Englisch, Französisch, Italienisch, Russisch, Spanisch) zum Thema Kulturgüter enthält (vgl. die didaktische Anwendung in Ballestracci/Bufagni/Flinz 2020).

Eher für Lehrende gedacht, die sich über mögliche Fehler in spezifischen Erwerbsphasen des Deutschen als Fremdsprache informieren möchten oder nach Indikatoren für die Niveaueinstufung von LernerInnen suchen (vgl. Hirschmann 2015 zu Modifikatoren), sind Lernerkorpora²⁵:

¹⁸ Computer-mediated communication Korpora (CMC-Korpora), Historische Korpora, L2-Lernerkorpora, literarische Korpora, manuell annotierte Korpora, Multimodale Korpora, Parallelkorpora, Zeitschriften-Korpora, Parlamentarische Korpora, Referenzkorpora, Spoken Korpora etc.

¹⁹ Neben den bereits erwähnten Korpora, dem *Dortmunder Chat Korpus*, den *Wikipedia Korpora* in DeReKo und dem *DWDS Blog Korpus* sind weitere enthalten: Das *Monitor Corpus of tweets from Austrian users* (Deutsch und Englisch, vgl. Barbaresi 2016), das *SMS-Corpus sms4science* (Schweizer Deutsch, Deutsch, Französisch, Italienisch etc. vgl. Dürscheid/Stark 2011); das *WhatsApp Korpus What's up, Switzerland?* (Schweizer Deutsch, Deutsch, Englisch, Italienisch etc.; vgl. Ueberwasser/Stark 2017).

²⁰ Vgl. <https://www.clarin.eu/resource-families/parallel-corpora>.

²¹ Vgl. <http://www.statmt.org/euoparl/>

²² Vgl. <https://intercorp.korpus.cz/?lang=de>

²³ Zum Aufbau und Benutzung von Vergleichskorpora in *Sketch Engine* für DaF vgl. Flinz (2020).

²⁴ Meistens werden Vergleichskorpora nur für spezifische Zwecke (lexikographische Zwecke, Diskursanalysen, kontrastive Analysen) erstellt und werden entweder gar nicht oder nur auf Anfrage zu Verfügung gestellt.

²⁵ Zum Thema *Lernerkorpora* wird Ende 2022 eine Themenausgabe der *KorDaF* erscheinen.

- das *Commented Learner Corpus Academic Writing / Kommentiertes Lernerkorpus akademisches Schreiben* von Dagmar Knorr (vgl. Aufsatz in diesem Themenheft), das authentische, deutschsprachige akademische Texten von Studierenden beinhaltet;
- *MERLIN Written Learner Corpus für Czech, German, Italian*. Es handelt sich um ein Korpus der geschriebenen Sprache, das zum Ziel der Illustration des *Common European Framework of Reference for Languages* (CEFR) realisiert worden ist. Es enthält 2.286 Texte von Deutsch-, Italienisch- und Tschechischlernenden, die im Rahmen standardisierter Sprachtests (A1-C1) verfasst wurden (vgl. Boyd/Hana/Lionel/Wisniewski/Abel/Schöne/Štindlová/Vettori 2014);
- *Falko. Ein fehlerannotiertes Lernerkorpus des Deutschen als Fremdsprache*. Es ist ein frei zugängliches fehlerannotiertes Lernerkorpus des Deutschen als Fremdsprache, das in der Annis-Umgebung untersucht werden kann. Es besteht aus mehreren Einzelkorpora: dem Essay-Kernkorpus, dem Essay-Vergleichskorpus mit Daten von MuttersprachlerInnen, dem WHiG-Korpus, dem Kobalt-DaF-Korpus, dem KanDel-Korpus, dem Zusammenfassungskorpus und dem Georgetown-Longitudinalkorpus).

Auch das *World Wide Web* kann als Korpus mit fremdsprachendidaktischer Zielsetzung benutzt werden (vgl. die Vermittlung sprachlicher Charakteristika unterschiedlicher Genres). Probleme können jedoch die Aussagekraft der Daten bereiten, da nicht immer deutlich ist, ob die Texte von MuttersprachlerInnen verfasst worden sind (vgl. Mukherjee 2009: 60); hinzu kommen weitere organisatorische und methodische Probleme (vgl. Baroni/Bernardini 2006).

4 Ansätze in der Korpuslinguistik und DDL

In der Korpuslinguistik²⁶ können Untersuchungen mit unterschiedlichen Ansätzen durchgeführt werden. Lemnitzer/Zinsmeister (2015: 34) unterscheiden auf der Basis der Dichotomien *Empirismus/Rationalismus* und *Deduktion/Induktion* drei Ansätze: korpusbasiert quantitativ; korpusbasiert, quantitativ und qualitativ; korpusgestützt²⁷ (vgl. Abbildung 23):

²⁶ Es soll nur darauf hingewiesen werden, dass der Status der Korpuslinguistik als Methode oder als Disziplin umstritten war bzw. ist (vgl. dazu Mukherjee 2009: 51–53). In diesem Aufsatz wird in Anlehnung an Lemnitzer/Zinsmeister (2015: 14) die Auffassung vertreten, dass Korpuslinguistik eine wissenschaftliche Disziplin ist.

²⁷ Zu diesem Punkt ist eine terminologische Klärung notwendig, da das deutsche Wort *korpusbasiert* oft mit dem englischen *corpus-based* in Verbindung gebracht wird. *Korpusbasiert* entspricht hingegen der Bezeichnung *corpus driven* von Tognini-Bonelli (2001: 84), d.h. das Korpusdaten genutzt werden, um Hypothesen zu erstellen. *Korpusgestützt* entspricht hingegen

	Korpusbasiert quantitativ	Korpusbasiert, quantitativ und qualitativ	Korpusgestützt
Ansätze	Latent-semantische Analyse N-Gramm Analyse	Koselektion und Kollokation semantische Prosodie	Wortstellungsphänomene
Theoretischer Rahmen	(nicht anwendbar)	Kontextualismus (Firth)	Strukturalismus (Saussure) / Generative Grammatik (Chomsky)
Erkenntnistheoretischer Ansatz	Extrem empiristisch	Gemäßigt empiristisch	Rationalistisch
Personen	Landauer – Jelinek	Sinclair, Teubert, Heringer	Fillmore, Arts, Oostdijk, Reis, Meurers
Eingabe	Korpus in Rohform	Korpus in Rohform	Linguistisch annotiertes Korpus oder Belegsammlung
Ausgabe	Text-Term Matrizen – N-Gramme mit Frequenzen	Kollokator-Kollokanten-Paare mit Kennziffern	Belegsätze
Interpretation	Keine	Ja, von den Belegen ausgehend	Ja, von den theoretischen Aussagen ausgehend
Primäre linguistische Domäne	Statistische Sprachmodelle	Semantik	Syntax
Anwendungsbereich	Informationserschließung, Verarbeitung gesprochener Sprache	Lexikographie, Fremdsprachunterricht, Übersetzungswissenschaft	Theoretische Linguistik

Abb. 23: Ansätze nach Lemnitzer/Zinsmeister (2015: 34)

Beim korpusbasierten quantitativen Ansatz werden quantitative Daten von rohen (also nicht annotierten) Korpora und von annotierten Korpora²⁸ extrahiert. Mögliche quantitative Analysen sind:

1. Extrahierung von Häufigkeitslisten eines Untersuchungskorpus (absolute Häufigkeit) (wie mit *Wordlist* in *Sketch Engine*);
2. Untersuchungen zur Häufigkeit eines Wortes oder Lemmas (absolute Häufigkeit) (wie mit *Concordance* in *Sketch Engine*; mit der Suchanfrage in *COSMAS II*; mit der *DDC* und mit der *Wortverlaufskurve* im DWDS-Wortinformationssystem; mit der Tokensuche in der *DGD*);
3. Extrahierung von Häufigkeitslisten eines Untersuchungskorpus auf der Basis eines Kontrollkorpus (relative Häufigkeit);

der Bezeichnung *corpus based* (vgl. Tognini-Bonelli 2001: 65), d.h. das Korpus wird genutzt, um Hypothesen zu verifizieren.

²⁸ Wir unterscheiden uns in dieser Hinsicht von Lemnitzer/Zinsmeister (2015: 37), die nur von Rohdaten in diesem Ansatz sprechen. Quantitative Untersuchungen können auch in annotierten Korpora (wie dem obengenannte *German Web 2013*) getätigt werden, während bei vielen großen Rohkorpora (wie *DeReKo*) einige quantitative Aufgaben nicht autonom durchgeführt werden können (Erstellung von Häufigkeitslisten, Keywordslisten etc.).

4. Extrahierung von Keyword-Listen eines Untersuchungskorpus auf der Basis eines Referenzkorpus (unterschiedliche statistische Assoziationsmaßen). Diese Listen werden normalerweise als Ranglisten erstellt (wie mit *Keywords in Sketch Engine*);
5. Die Distribution eines Wortes in einem Untersuchungskorpus;
6. Extrahierung von Kollokationen, syntagmatische Muster, N-Gramme (wie mit der Kookkurrenzanalyse in *COSMAS II*; wie mit der *CCDB*; wie mit dem *Wortprofil* und *DiaCollo* im DWDS-Wortinformationssystem; wie mit *Word Sketch* und *N-Grams* in *Sketch Engine*);
7. Semantische Ähnlichkeiten oder Unterschiede zwischen zwei Wörtern auf der Basis des Kollokationsprofils (wie mit *DDC* im DWDS-Wortinformationssystem; mit *Thesaurus* und *Sketch Difference* in *Sketch Engine*).

Die qualitative Interpretation dieser Daten ist in diesem Ansatz nicht notwendig.

Der korpusbasierte quantitativ-qualitative Ansatz ist dem gerade vorgestellten sehr ähnlich, denn die Hauptquelle der linguistischen Untersuchungen ist weiterhin die Beobachtung des Sprachgebrauchs (vgl. Lemnitzer/Zinsmeister 2015: 37), aber die extrahierten Daten bleiben nicht uninterpretiert. Sie werden meistens auf der Basis von grammatischen Kategorien interpretiert. Im korpusgestützten Ansatz werden hingegen Korpusdaten nur als Quelle der Evidenz genutzt. Korpora werden nicht als Ganzes herangezogen, sondern zur Suche nach lexikalischen oder grammatischen Elementen sowie nach syntaktischen Konstruktionen, um Hypothesen zu verifizieren oder zu falsifizieren.

Für den Fremdsprachenunterricht²⁹ wird von Lemnitzer/Zinsmeister (2015: 37) vor allem der korpusbasierte quantitativ-qualitative Ansatz vorgeschlagen: Die Daten werden aus Korpora abgeleitet, werden jedoch interpretiert. Wir stimmen hier zu, sind aber auch der Meinung, dass der korpusgestützte Ansatz von großer Bedeutung für den Fremdsprachenunterricht sein kann, denn die Suche nach einem Wort und die Analyse seines Kontextes können auch induktive Beobachtungen leiten.

Korpuslinguistische Untersuchungen können in der DDL-Didaktik sowohl indirekt als auch direkt genutzt werden. Im ersten Fall sind sie Anlass für die Erstellung von Lernerwörterbüchern, DaF-Grammatiken, DaF-Lehrbücher, Kursmaterialien oder werden für die Unterrichtsvorbereitung verwendet (vgl. u.a. Moellering 2004 zu Modalpartikeln und Jones 2000 zu Präpositionen). Im zweiten Fall werden Korpora direkt im Unterricht benutzt: Sie dienen als Quelle für authentische Beispiele und bieten den Lernenden einen ersten Kontakt zur natürlichen Sprachverwendung (vgl.

²⁹ Neben dem Fremdsprachenunterricht hat Korpuslinguistik weitere Anwendungsbereiche, wie u.a. die computergestützte Sprachsynthese und -verarbeitung, die Spracherwerbsforschung, die forensische Linguistik (vgl. Meukherjee 2009).

Lemnitzer/Zinsmeister 2015: 189). Diese Anwendung wird auch „On Stage Approach“ bzw. „Auf der Bühne-Ansatz“ genannt (vgl. Aston 2000) und ist noch unterbelichtet (vgl. Corino 2019: 272; Römer 2009; 2010; Tribble 2015). Deswegen möchten wir einige Beschreibungsebenen skizzieren, um zu zeigen, wie facettenreich die Benutzung von Korpora im Unterricht sein kann. Sprachanalyse und natürliche Verwendung werden vereinheitlicht, so dass Korpora potenziell für alle Sprachebenen benutzt werden können:

- Orthographie: u.a. Verwendung des Bindestrichs (Zusammenschreibung oder Schreibung mit Bindestrich); Binnengroßschreibung und Getrennschreibung von Komposita; Untersuchung von graphostilistischen Elementen (vgl. Lemnitzer/Zinsmeister 2015: 157–160);
- Phonethische und phonologische Ebene: u.a. Aussprachevarianten (vgl. Kaiser/Schedl in diesem Themenheft), die Unterscheidung zwischen Standardvarietäten und Dialekt; Fragen nach der Häufigkeit einzelner Phoneme und Allophone, sowie nach ihren Verbindungen; die kontext- und kotextsensitive Reduzierbarkeit von Vokalen und Konsonanten; die Richtung und Art der Assimilation (vgl. Fandrych/Tschirner 2007: 197);
- Morphosyntaktische und lexikalische Ebene: u.a. Aspekte wie Häufigkeit und Gebrauch bestimmter grammatischer Phänomene; Häufigkeit und Gebrauch einzelner Wortbausteine (Wortstämme, Affixe, Flexive) (vgl. Gredel in dieser Themenausgabe); Produktivität von Wortbildungselementen, da die Vermittlung ihrer morphologischen und semantischen Regularität eine wichtige Rolle spielt, da Lernende Wörter diesen Bildungstyps nicht in Wörterbüchern finden werden (vgl. Lemnitzer/Zinsmeister 2015: 165); Häufigkeit und Gebrauch verschiedener Flexionsformen eines Lexems; Häufigkeit der verschiedenen Verbformen (Tempora); das Verhältnis zwischen Grammatik und Lexik (fließende Übergänge, gegenseitige Beeinflussung wie bei der Valenz, bei negativen Polaritätselementen³⁰, Wechselpräpositionen³¹ und bei bestimmter Formulierungsroutinen³²); Vergleich Ziel- und Ausgangssprache; textsortenspezifische Häufigkeit bestimmter grammatischer Strukturen; Unterschiede zwischen Mündlichkeit und Schriftlichkeit; Wechselwirkungen zwischen Intonation und

³⁰ Vgl. Ausdrücke wie *unversucht lassen*, *ganz geheuer* etc., die nur in negierten Satzzusammenhängen oder in Fragesätzen vorkommen (vgl. Fandrych/Tschirner 2007: 197).

³¹ Der Kasus der Wechselpräpositionen wird nur selten aufgrund der Unterscheidung zwischen Ort und Ziel gewählt, sondern Präpositionen scheinen Affinität zu einem Kasus zu haben: die Präposition über hat eine Affinität von 99% zum Akkusativ, hinter 80% zum Dativ (vgl. Fandrych/Tschirner 2007: 197).

³² Bestimmte Formulierungen stehen in Verbindung mit spezifischen Sprachhandlungen: Textkommentierungen in der Wissenschaftssprache mit Modalverben, Passiv, spezifische Kollokationen.

Syntax; Vergleich der Satzlänge; Häufigkeit und Bedeutung von bestimmten Kollokationen, Phraseme, Konstruktionen (vgl. Gredel und Krekeler in dieser Themenausgabe); Vorfeldbesetzung; Distribution von Nominalverbgefügen; Gebrauch von Nebensätzen; Neologismen; Okkasionalismen; Anglizismen; Partikeln und insbesondere Modalpartikeln (vgl. Lemnitzer/Zinsmeister 2015: 183), Gradpartikeln (Nederstigt 2003), Intensitätspartikeln;

- Textlinguistik;
- Pragmatik: u.a. Gesprächspartikeln; sprachliche Handlungen der Kategorie „Vorschlag/Angebot“ (vgl. Kaiser/Schedl in diesem Themenheft), Komplimente (vgl. Ravetto/Castagneto in dieser Themenausgabe); das pragmatische Phänomen der Sozialdeixis (Siezen und Duzen) (vgl. Gredel in dieser Themenausgabe);
- Stilistik: u.a. stilistische Abweichungen von einer Norm, Umgangssprache; Jugendsprache;
- Soziolinguistik: Gender Studies; Inklusion vs. Exklusion; Stereotypen und Vorurteile.

Neben der Auswahl der linguistischen Sprachebene sind bei der Benutzung von Korpora im Unterricht zwei zentrale Aspekte zu berücksichtigen: das Niveau, auf dem sich die Lernenden befinden bzw. das angestrebt wird, und die Fertigkeit, die vermittelt werden soll (vgl. Tschirner 2003; Perkuhn et al. 2012: 69).

5 Vorstellung der Beiträge

Die vorliegende Themenausgabe soll ein erster Schritt in Richtung Verfestigung und Ausbreitung des DDL in DaF und DaZ sein³³. Unterschiedliche Korpusstypen kommen zur Anwendung und werden diskutiert: mündliche Korpora, wie *FOLK* (Beitrag von Julia Kaiser und Evi Schedl); das Korpus *Compliment Corpus – Co.Cor* (Beitrag von Miriam Ravetto und Marina Castagneto), *CMC Korpora* (Beitrag von Eva Gredel); Lernerkorpora, wie das Korpus *Commented Learner Corpus Academic Writing/Kommentiertes Lernendenkorpus akademisches Schreiben* (Beitrag von Dagmar Knorr) und schriftliche Korpora, wie *Deutscher Wortschatz* (<https://corpora.uni-leipzig.de/>), das *Deutsche Referenzkorpus* und ein kleines Spezialkorpus von 68 Dissertationen aus der Betriebswirtschaftslehre (Beitrag von Christian Krekeler).

In dem Aufsatz „Das Forschungs- und Lehrkorpus Gesprochenes Deutsch als Ressource für den handlungsorientierten DaF-Unterricht – Potentiale und Herausforderungen für die Nutzung im handlungsorientierten DaF- (und eventuell auch DaZ-) Unterricht fruchtbar gemacht zu werden“ zeigen Julia Kaiser und Evi Schedl, wie das Korpus *FOLK* als Ressource für den Unterricht im Fach Deutsch als Fremdsprache benutzt werden kann. Der inhaltliche Fokus liegt auf dem Bereich sprachlicher Handlungen und ihrer Realisierung in Form von Modalverbformaten. Nach einem Überblick über die konversationsanalytische Forschung im Bereich der Sprachlehre und nach Absteckung des theoretischen Rahmens des Beitrages (Mündlichkeit und Interaktion in der DaF-Didaktik) folgen Erläuterungen zur Auswahl und Zusammenstellung eines *ad hoc* *FOLK*-Subkorpus. Das Subkorpus wird beschrieben, sowie der Workflow zu seiner inhaltlichen Aufbereitung durch Annotationen. Sprachdidaktische Potenziale bei der Nutzung der annotierten Gesprächsausschnitte für die Modalverbformate *Sollen wir X?* und *Wollen wir X?* für sprachliche Handlungen der Kategorie „Vorschlag/Angebot“ werden vorgestellt und zur Diskussion gestellt.

Die Beschreibung des Korpus *Compliment Corpus* (*Co.Cor*) und seine didaktische Anwendung für die Entwicklung von pragmatischen Kompetenzen sind im Zentrum des Beitrages von Miriam Ravetto und Marina Castagneto, der den Titel „Zum Erwerb pragmatischer Kompetenz im Fremdsprachenunterricht: das Korpus

³³ Ein weiter Schritt ist die Gründung einer digitalen Zeitschrift zu diesem Thema, die an der Technischen Universität Darmstadt gehostet sein wird; sie wird zweimal jährlich erscheinen. Eine aktive Mitwirkung, sowohl als VerfasserInnen von wissenschaftlichen Aufsätzen, Praxisberichten oder Präsentationen von Korpora und Projekten als auch als GutachterInnen soll mit dieser Themenausgabe angestrebt werden. Bei Interesse kontaktieren Sie die Herausgeberinnen (Carolina Flinz – carolina.flinz@unimi.it; Britta Hufeisen – britta.hufeisen@tu-darmstadt.de) oder die Redaktion (Ruth Mell – ruth_maria.mell@tu-darmstadt.de).

Co.Cor (Compliment Corpus)“ trägt. Das Korpus Co.Cor ist eine Datenbank, die aus Audio- und Videoaufzeichnungen authentischer (semi-)spontaner Gespräche besteht, in denen der Sprechakt des Kompliments realisiert wird. Es handelt sich um ein multilinguales Korpus, das außer Italienisch und Deutsch weitere Sprachen (Russisch und Albanisch) beinhaltet; die Aufzeichnungen sind nach den AVIP-Konventionen transkribiert. Nach Erläuterung des theoretischen Hintergrundes (die pragmatische Kompetenz und ihre Rolle beim Fremd- bzw. Zweitspracherwerb) wird Co.Cor beschrieben. Anschließend werden die Sprachhandlungen des Kompliments und der Reaktion auf Komplimente behandelt, sowie die möglichen Antworttypen klassifiziert. Die Ergebnisse eines empirischen Versuchs werden abschließend diskutiert, und es wird gezeigt, welche Übungsaktivitäten zum Ziel der Entwicklung von pragmatischen Kompetenzen mit dem Korpus getätigt werden können. Zu diesem Zweck sollen LernerInnen authentische Interaktionen und Sprechakte in der Fremdsprache betrachten, untersuchen, durch Rollenspiele üben und sie mit ähnlichen bzw. analogen Situationen in der Erstsprache vergleichen.

Im Beitrag „CMC-Korpora und Digital Literacy in der Fremdsprachendidaktik: Relevanz, Potentiale und didaktische Szenarien“ von Eva Gredel geht es um die Entwicklung von *Digital Literacy* in der Fremdsprachendidaktik auf der Basis von CMC-Korpora (*Computer Mediated Communication*). Für die obengenannte Integration (*Digital Literacy*/CMC-Korpora) werden die Argumente der Authentizität sowie der Erweiterung des Spektrums schriftsprachlicher Handlungsmöglichkeiten angeführt. Nach der Skizzierung des Forschungsstandes (*Digital Literacy* als Kompetenz und die Beschäftigung mit CMC-Korpora im DaF-/DaZ-Bereich) wird ein Überblick über verfügbare CMC-Korpora gegeben. Fünf didaktische Szenarien am Beispiel der Wikipedia-Korpora des IDS werden anschließend vorgestellt und diskutiert. Sprachliche Phänomene auf ganz verschiedenen Ebenen werden dabei fokussiert: das morphologisch-syntaktische Phänomen der Inflektive bzw. Asteriskausdrücke als Marker für den interaktionsorientierten Schreibstil; Bedeutung und spezifische Muster von lexikalischen Einheiten wie das Lexem *Baustein*; das pragmatische Phänomen der Sozialdeixis (Siezen und Duzen); die Nutzung bestimmter Schlag- und Schlüsselwörter in europapolitischen Diskursen.

In dem Beitrag „Zwischen Forscher-, Verfasser- und Erzähler-Ich. Eine korpuslinguistische Studie zur Konstruktion von Selbstreferenz und ihre Einsatzmöglichkeit in der Schreibberatungsausbildung“ von Dagmar Knorr geht es darum, wissenschaftlich schreiben zu lernen, was ein langer und komplexer Prozess ist. Schreibzentren haben die Rolle, Studierende durch Schreibberatung dabei zu unterstützen, und studentische SchreibberaterInnen werden auch zu diesem Zweck geschult. Thema dieses Aufsatzes ist der Gebrauch von *ich*, ein sprachliches Mittel für die Konstruktion von Selbstreferenz, was auch ein Problem für studentische

SchreibberaterInnen darstellt, die häufig unsicher sind, wie *ich* gebraucht werden darf. Die präsentierte Studie zeigt, wie korpuslinguistische Methoden eingesetzt werden können, um hierzu einen Beitrag zu leisten. Datengrundlage ist das von der Autorin erstellte Lernendenkorpus *Commented Learner Corpus Academic Writing/ Kommentiertes Lernendenkorpus akademisches Schreiben*. Ausgangspunkt ist die Typisierung von Steinhoff (2007), die zwischen Verfasser-, Forscher- und Erzähler-Ich unterscheidet. Für die Annotation wird den drei Ich-Typen eine zusätzliche *-Kategorie hinzugefügt. Geprüft wurde, inwieweit studentische Schreibberaterinnen, die sich im Rahmen eines Workshops mit dem Artikel von Steinhoff auseinandergesetzt haben, in der Lage sind, die Typisierung auf das Lernerkorpus anzuwenden. Die Grundannahme ist, dass die Annotation durch eine Zielhypothese beeinflusst wird. Im Rahmen einer exemplarischen qualitativen Analyse, die Zielthesen der AnnotatorInnen rekonstruiert, wird hervorgehoben, dass Zielthesen (text)diagnostisch oder auf eine Anschlusshandlung hin orientiert sein können. Diese Orientierung beeinflusst die Interpretation und Typisierung von Ich-Vorkommen. Die Autorin veranschaulicht, wie die Arbeit mit Korpora vielfältige Möglichkeiten in der Schreibwissenschaft eröffnen kann: Die Ergebnisse können genutzt werden, um Konzepte für eine konstruktive und sprachensensible Textkommentierung und für die Erstellung von Selbstlernmaterialien zu entwickeln, die Lernende für die Selbstreflexion nutzen können, oder aber auch, um Lehrende für die Problematik der Konstruktion von Verfasserreferenz zu sensibilisieren.

Im praktisch orientierten Aufsatz „Korpustraining und datengestütztes Lernen im Wirtschaftsdeutsch-Unterricht“ von Christian Krekeler wird am Beispiel einer Unterrichtseinheit für Wirtschaftsdeutsch der Einsatz von DDL erläutert. Es wurden die Korpora *Deutscher Wortschatz* (<https://corpora.uni-leipzig.de/>), das *Deutsche Referenzkorpus* und ein kleines Spezialkorpus (68 Dissertationen aus der Betriebswirtschaftslehre) benutzt, während als Tools *AntConc*, *WordSmith* (Lexical Analysis Software 2020) und das Programm *QuAXDaF* eingesetzt wurden. Chancen und Probleme des Einsatzes von DDL werden diskutiert und Vorschläge für seine Verwendung in Fachsprachenunterricht vorgestellt. Vorerst wird der indirekte Gebrauch des DDLs (in diesem Fall für die Unterrichtsplanung) thematisiert, anschließend wird das Praxisbeispiel mit den unterschiedlichen Übungen veranschaulicht. Der Fokus liegt auf sprachlichen Mustern, auch mit Thematisierung der Fragen nach der Korrektheit bzw. der Angemessenheit.

Wie diesem synthetischen Überblick über die einzelnen Beiträge entnommen werden kann, können Korpora sowohl von den Lehrenden als auch von den Lernenden verwendet werden. Im Fokus können sowohl rezeptive als auch produktive Fertigkeiten sein. Indirekt können die Korpora für Unterrichtsplanung (s. Kaiser/Schedl, Krekeler) benutzt werden; direkt können sie im Unterricht eingesetzt werden, um

spezifische Kompetenzen zu entwickeln: u.a. phonetisch-phonologische (s. Kaiser/Schedl), morphologisch-syntaktische (s. Gredel), lexikalische (s. Gredel, Krekeler) und pragmatische (s. Kaiser/Schedl, Ravetto/Castagneto, Gredel). Korpora können jedoch auch für die Korrektur (s. Krekeler) und Selbstreflektion (s. Knorr) herangezogen werden. Außerdem können Lernerkorpora eingesetzt werden, um sprachliche Mittel, wie den Gebrauch von *ich* für die Konstruktion von Selbstreferenz, zu untersuchen und in der Schreibwissenschaft Anlass zur Reflektion über konstruktive und sprachensensible Textkommentierung sein sowie die Erstellung von adäquaten Selbstlernmaterialien stützen.

Literatur

- Aston, Guy (2000): Learning English with the British National Corpus. Paper presented at *6th Jornada de Corpus Linguistics. Barcelona*. <http://www.sslmit.unibo.it/~guy/barc.htm> (01.03.21).
- Ballestracci, Sabrina; Buffagni, Claudia & Flinz, Carolina (2020): *Das deutsche LBC-Korpus: Zusammenstellung und Anwendung*. In *Vrb.*, <http://corpora.lessicobeniculturali.net/de/> (01.03.21).
- Barbaresi, Adrien (2016): Collection and Indexing of Tweets with a Geographical Focus. In: *Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, May 2016, Portorož, Slovenia, 24–27, <https://hal.archives-ouvertes.fr/hal-01323274v2> (01.03.21).
- Baroni, Marco & Bernardini, Silvia (eds) (2006): *WaCky! Working papers on the web as corpus*. Bologna: Gedit.
- Boulton, Alex (2009): Data-driven learning: reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics* 35: 1, 81–106.
- Boulton, Alex (2010): Consultation de corpus et styles d'apprentissage. *Cahiers de l'APLIUT* 29:1, 98–115.
- Boulton, Alex (2011): Language awareness and medium-term benefits of corpus consultation. In: Gimeno Sanz, Ana (ed.): *New trends in computer-assisted language learning: working together*. Madrid: Macmillan ELT, 39–46.
- Boulton, Alex (2020): Data-Driven Learning for Younger Learners: Obstacles and Optimism. In: Crosthwaite Peter (ed.): *Data-Driven Learning for the Next Generation*. New York: Routledge, XIV–XX.
- Boulton, Alex & Tyne, Henry (2013): Corpus linguistics and data-driven learning: A critical overview. *Bulletin suisse de Linguistique appliquée*. Neuchâtel: Institut de linguistique de l'Université. Rôle des pratiques langagières dans la constitution des espaces sociaux pluriels d'aujourd'hui Actes du colloque VALS-ASLA 2012 (Lausanne, 1-3 février 2012), 97–118, <https://hal.archives-ouvertes.fr/hal-01850687/document> (01.03.2021).
- Boyd, Adriane; Hana, Jirka; Lionel, Nicolas; Meurers, Detmar; Wisniewski, Katrin; Abel, Andrea; Schöne, Karin; Štindlová, Barbora & Vettori, Chiara (2014): The MERLIN corpus: Learner language and the CEFR. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. http://www.lrec-conf.org/proceedings/lrec2014/pdf/606_Paper.pdf (01.03.21).
- Braun, Sabine (2007): Integrating Corpus Work into Secondary Education: from Data-Driven Learning to Needs-Driven Corpora. *ReCALL* 19:3, 307–328, <https://doi.org/10.1017/S0958344007000535> (01.03.2021).
- Corino, Elisa (2014): Didattica delle lingue corpus-based. *EL.LE* 3: 2, 231–257.
- Corino, Elisa (2019): Data-driven Learning: tra lingue straniere e CLIL, tra ricerca e didattica. *EL.LE* 8: 2, 271–286.

- Corino, Elisa & Onesti, Cristina (2019): Data-Driven Learning: A Scaffolding Methodology for CLIL and LSP Teaching and Learning. *Frontiers in Education*, 4, art. 7, 1–12. <https://doi.org/10.3389/feduc.2019.00007> (01.03.2021).
- Deppermann, Arnulf & Hartung, Martin (2011): Was gehört in ein nationales Gesprächskorpus? Kriterien, Probleme und Prioritäten der Stratifikation des 'Forschungs- und Lehrkorpus Gesprochenes Deutsch' (FOLK) am Institut für Deutsche Sprache (Mannheim). In: Felder, Ekkehard; Müller, Marcus & Vogel, Friedemann (Hrsg.): *Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen*. Berlin/New York: de Gruyter, 414–450.
- Deppermann, Arnulf & Schmidt Thomas (2014): Gesprächsdatenbanken als methodisches Instrument der Interaktionalen Linguistik - Eine exemplarische Untersuchung auf Basis des Korpus FOLK in der Datenbank für Gesprochenes Deutsch (DGD2). *Mitteilungen des Deutschen Germanistenverbandes* 61: 1, 4–17.
- Dietz, Gunther (2021): Authentische Hörmaterialien im DaZ-/DaF-Unterricht – Korpora gesprochener Sprache als Fundgrube für die Erstellung von Mikro-Höraufgaben. *Korpora DaF* 1/2021. In Vrb.
- Dürscheid, Christa & Stark, Elisabeth (2011): sms4science: An international corpus-based texting project and the specific challenges for multilingual Switzerland. In: Thurlow, Crispin & Mroczek, Kristine (Hg.): *Digital Discourse: Language in the New Media*. Oxford: Oxford University Press, 299–320.
- Fandrych, Christian; Meißner, Cordula & Wallner, Franziska (2018): Das Potenzial mündlicher Korpora für die Sprachdidaktik: Das Beispiel GeWiss. In: *Deutsch als Fremdsprache* 55:1, 3–13.
- Fandrych, Christian; Meißner, Cordula & Wallner, Franziska (2021): Gesprochene Sprache im DaF/DaZ-Unterricht: Zugänge zu authentischem sprachlichem Handeln mit ZuMult. In: Hinzmann Friederike; Storz Coretta & Thielmann, Winfried (Hrsg.): *Beiträge der FaDaF-Jahrestagung an der TU Chemnitz 2019*. Göttingen: Göttinger Universitätsverlag (im Druck).
- Fandrych, Christian & Tschirner, Erwin (2007): Korpuslinguistik und Deutsch als Fremdsprache. Ein Perspektivenwechsel. *Deutsch als Fremdsprache* 4, 195–204.
- Flinz, Carolina (2020): Vergleichbare Spezialkorpora für den Tourismus. Eine Chance für den Fachsprachenunterricht. In: Hepp, Marianne & Salzmann, Katharina (Hrsg.): *Sprachvergleich in der mehrsprachig orientierten DaF-Didaktik. Theorie und Praxis*. Roma: Istituto Italiano di Studi Germanici, 133–151.
- 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, 5–82.

- Flinz, Carolina & Katelhön, Peggy (2019): Corpora nella didattica del tedesco come lingua straniera. Proposte per l'insegnamento del linguaggio specialistico del turismo. *EL.LE* 8: 2, 323–348.
- Flinz, Carolina & Perkuhn, 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 EURALEX International Congress: Lexicography in Global Context*. Ljubljana University Press, Faculty of Arts: Ljubljana, 959–967.
- Flowerdew, Lynne (2015): *Data-Driven Learning and Language Learning Theories: Whither the Twain Shall Meet*. In: Lenko-Szymanska, Agnieszka & Boulton, Alex (eds): *Multiple Affordances of Language Corpora for Data-driven Learning*. Amsterdam: John Benjamins, 15–36.
- Gabrielatos, Costas (2005): Corpora and language teaching: just a fling or wedding bells? *Teaching English as a Second Language – Electronic Journal* 8: 4, 1–35, <http://tesl-ej.org/ej32/a1.html> (01.03.2021).
- Gilquin, Gaëtanelle & Granger, Sylviane (2010): How can data-driven learning be used in language teaching? In: O'Keeffe, Anne & McCarthy, Michael (eds.): *Routledge handbook of corpus linguistics*. London: Routledge, 359–370.
- Hirschmann, Hagen (2015): *Modifikatoren im Deutschen. Ihre Klassifizierung und varietätenspezifische Verwendung*. Tübingen: Stauffenburg.
- Hirschmann, Hagen (2019): *Korpuslinguistik. Eine Einführung*. Stuttgart: Metzler.
- Hoey, Michael (2005): *Lexical priming: a new theory of words and language*. London: Routledge.
- Imo, Wolfgang & Weidner, Beate (2018): Mündliche Korpora im DaF- und DaZ-Unterricht. In: Kupietz, Marc & Schmidt, Thomas (Hrsg.): *Korpuslinguistik*. Berlin: de Gruyter, 231–251.
- Johansson, Stig (2009): Some Thoughts on Corpora and Second-Language Acquisition. In: Aijmer, Karin (ed.): *Corpora and Language Teaching*. Amsterdam: John Benjamins, 33–44.
- Johns, Tim (1988): Whence and whither classroom concordancing? In: Bongaerts, Theo; De Haan, Pieter; Lobbe, Sylvia & Wekker, Herman (eds.): *Computer applications in language learning*. Dordrecht: Foris, 9–27.
- Johns, Tim (1997): Contexts: The Background, Development and Trialling of a Concordance-Based CALL program. In: Wichmann, Anne; Fligelstone, Steven & McEnery, Tony (eds): *Teaching and Language Corpora*. Harlow: Addison Wesley Longman, 100–115.
- Jones, Randall L. (2000): A corpus-based study of German accusative/dative preposition. In: Dodd, Bill (Hrsg.): *Working with German corpora*. Birmingham: Birmingham University Press, 116–142.

- Kaiser, Julia (2018): Zur Stratifikation des FOLK-Korpus: Konzeption und Strategien. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 19, 515–552.
- Kirschner, Paul; Sweller, John & Clark, Richard (2006): Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and Inquiry-Based teaching. *Educational Psychologist* 41: 2, 75–86.
- Koch, Peter & Oesterreicher, Wulf (1994): Schriftlichkeit und Sprache. In: Günther, Hartmut und Ludwig, Otto (Hrsg.): *Schrift und Schriftlichkeit*. Berlin/New York: de Gruyter, 587–604.
- Kupietz, Marc; Lungen Harald; Kamocki Paweł & Witt Andreas (2018): The German Reference Corpus DeReKo: New Developments – New Opportunities. In: Calzolari, Nicoletta et al. (eds): *Proceedings of The Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki: European Language Resources Association (ELRA), 4353–4360.
- Kupietz, Marc & Schmidt Thomas (Hg.) (2018): *Korpuslinguistik*. Berlin/Boston: de Gruyter.
- Lemnitzer, Lothar & Zinsmeister, Heike (2006): *Korpuslinguistik. Eine Einführung*. Tübingen: Narr.
- Lemnitzer, Lothar & Zinsmeister, Heike (2010): *Korpuslinguistik. Eine Einführung*. 2. Auflage. Tübingen: Narr.
- Lemnitzer, Lothar & Zinsmeister, Heike (2015): *Korpuslinguistik. Eine Einführung*. 3. Auflage. Tübingen: Narr.
- Lüdeling Anke & Walter, Maik (2009): Korpuslinguistik für Deutsch als Fremdsprache. Sprachvermittlung und Spracherwerbsforschung. <https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/mitarbeiterinnen/anke/pdf/LuedelingWalterDaF.pdf> (01.03.2021). Stark erweiterte Fassung von Lüdeling, Anke & Maik, Walter: Korpuslinguistik, in Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch. In: Fandrych, Christian; Hufeisen, Britta; Krumm, Hans-J. & Riemer, Claudia (Hrsg.) (2010): *Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch*. 2. Halbband. Berlin, New York: De Gruyter Mouton, 315–322.
- Malloggi, Patrizio (2021): Zur korpusbasierten Vermittlung von Funktionsverbgefügen am Beispiel des DWDS im DaF- bzw. DaZ-Unterricht. *Korpora DaF* 1/2021. In Vrb.
- McEnery, Tony; Xiao, Richard & Tono Yukio (2006): *Corpus-Based Language Studies. An Advanced resource Book*. London: Routledge.
- Moellering, Martina (2004): *The Acquisition of German Modal Particles. A Corpus-based Approach*. Bern: Peter Lang.
- Moroni Manuela (2021): Zum Einsatz der Datenbank für Gesprochenes Deutsch in der universitären Unterrichtspraxis am Beispiel der Partikeln *doch* und *auch*. *Korpora DaF* 1 In Vrb.

- Mukherjee, Joybrato (2002): *Korpuslinguistik und Englischunterricht. Eine Einführung*. Frankfurt am Main: Peter Lang.
- Mukherjee, Joybrato (2006): Corpus linguistics and language pedagogy: the state of the art – and beyond. In: Braun, Sabine; Kohn, Kurt & Mukherjee, Joybrato (eds.): *Corpus technology and language pedagogy: new resources, new tools, new methods*. Frankfurt: Peter Lang, 5–24.
- Mukherjee, Joybrato (2008): Anglistische Korpuslinguistik und Fremdsprachenforschung. Entwicklungslinien und Perspektiven. *Zeitschrift für Fremdsprachenforschung* 19: 1, 31–60.
- Mukherjee, Joybrato (2009): *Anglistische Korpuslinguistik: Eine Einführung*. Berlin: Erich Schmidt.
- Nolting, Antje & Radtke, Nadja (2019): Korpusbasierte Lexikografie. Nutzung von Korpora und Analysewerkzeugen im Unterricht für Deutsch als Muttersprache und Fremdsprache. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 24:1, 107–126. <https://tujournals.ulb.tu-darmstadt.de/index.php/zif/article/view/953/952> (16.03.2021).
- Perkuhn, Rainer; Keibel, Holger, & Kupietz, Marc (2012): *Korpuslinguistik*. Paderborn: Fink.
- Nederstigt, Ulrike (2003): *Auch and noch in child and adult German*. [Studies in Language Acquisition, Band 23]. Berlin/New York: Mouton de Gruyter.
- Portmann-Tselikas, Paul (2001): Sprachaufmerksamkeit und Grammatiklernen. In: Portmann-Tselikas, Paul & Schmölzer-Eibinger, Sabine (Hrsg.): *Grammatik und Sprachaufmerksamkeit*. Innsbruck: Studienverlag, 9–48.
- Römer, Ute (2006): Pedagogical Applications of Corpora: Some Reflections on the Current Scope and a Wish List for Future Developments». *ZAA* 54: 2, 121–134.
- Römer, Ute (2008): Corpora and Language Teaching. In: Lüdeling, Anke & Kytö, Merja (eds): *Corpus Linguistics. An International Handbook*, vol. 1. Berlin; New York: Mouton de Gruyter, 112–31.
- Römer, Ute (2009): Corpus Research and Practice: What Help Do Teachers Need and What can We Offer? In: Aijmer, Karin (ed.): *Corpora and Language Teaching*. Amsterdam: John Benjamins, 83–98.
- Römer, Ute (2010): Using General and Specialized Corpora in English Language Teaching: Past, Present and Future. In: Campoy-Cubillo, Marí C.; Bellés-Fortuño, Begona & Gea-Valor, Lluisa (eds.): *Corpus-based approaches to English language teaching*. London: Continuum, 18–35.
- Rundell, Michael & Stock, Penny (1992): The Corpus Revolution. *English Today* 8: 4, 45–51.
- Scherer, Carmen (2006): *Korpuslinguistik*. Heidelberg: Winter.
- Scherer, Carmen (2014): *Korpuslinguistik*. 2. Aktualisierte Ausgabe. Heidelberg: Winter.

- Sieberg, Bernd (2016): Reaktive: Vorschlag für eine Erweiterung der Kategorie Responsive. In: Handwerker, Brigitte; Bäuerle, Rainer & Sieberg, Bernd (Hrsg.): *Gesprochene Fremdsprache Deutsch*. Baltmannsweiler: Schneider Hohengehren, 101–116.
- Sinclair, John (ed.) (1987): *Looking up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. London/Glasgow: Collins ELT.
- Sinclair, John (1996): *EAGLES Preliminary recommendations on Corpus Typology*. <http://www.ilc.cnr.it/EAGLES96/corpus/corpus.html> (01.03.21)
- Stefanowitsch, Anatol (2020): *Corpus linguistics. A guide to the methodology*. Berlin: Language Science Press.
- Steinhoff, Torsten (2007): Zum ich-Gebrauch in Wissenschaftstexten. *ZGL* 35: 2, 1–26.
- Steyer, Kathrin (2008): Kollokationen in deutschen Wörterbüchern und in der deutschen Wörterbuchforschung. *Lexicographica: International Annual for Lexicography* 24, 185–207.
- Swan, Michael (2005): Legislation by hypothesis: the case of task-based instruction. *Applied Linguistics* 26: 3, 376–401.
- Taylor, John (2012): *The mental corpus: how language is represented in the mind*. Oxford: Oxford University Press.
- Tognini-Bonelli, Elena (2001): *Corpus Linguistics at Work*. Amsterdam: Benjamins.
- Tribble, Christopher (2015): Teaching and Language Corpora: Perspectives from a Personal Journey. In: Lenko-Szymanska, Agnieszka & Boulton, Alex (eds): *Multiple Affordances of Language Corpora for Data-driven Learning*. Amsterdam: John Benjamins, 37–62.
- Tschirner, Erwin (2003): Skill, text, and register. Rethinking grammar in the IT age. In: Renandya, Willy (Hg.): *Methodology and Materials Design in Language Teaching. Current Perceptions and Practices and their Implications*. Singapore: SEAMEO Regional Language Centre, 70–87.
- Ueberwasser, Simone & Stark, Elisabeth (2017): What's up, Switzerland? A corpus-based research project in a multilingual country. *Linguistik Online* 84: 5. <https://doi.org/10.13092/lo.84.3849> (01.03.21).
- Varley, Steve (2009): I'll just look that up in the concordancer: integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning* 22: 2, 133–152.
- Wallner, Franziska (2013): Korpora im DaF-Unterricht - Potentiale und Perspektiven am Beispiel des DWDS. *Revista Nebrija de Lingüística Aplicada* 13. <https://www.nebrija.com/revista-linguistica/korpora-im-daf-unterricht-potentiale-und-perspektiven-am-beispiel-des-dwds.html> (01.03.2021).

Widdowson, Henry (2000): On the limitations of linguistics applied. *Applied Linguistics* 21: 1, 3–25.

Korpora

CLARIN - European Research Infrastructure for Language Resources and Technology. Resource families. <https://www.clarin.eu/resource-families>

Commented Learner Corpus Academic Writing / Kommentiertes Lernendenkorpus akademisches Schreiben. <http://hdl.handle.net/11022/0000-0006-CD41-A>

Das Deutsche Referenzkorpus – DeReKo. <https://cosmas2.ids-mannheim.de/cosmas2-web/>

DWDS. Der deutsche Wortschatz von 1600 bis heute. <https://www.dwds.de>

European Parliament Proceedings Parallel Corpus 1996–2011. <http://www.statmt.org/euoparl/>

Falko. Ein fehlerannotiertes Lernerkorpus des Deutschen als Fremdsprache. <https://korpling.german.hu-berlin.de/falko-suche/>

FOLK. Forschungs- u. Lehrkorpus für gesprochenes Deutsch. hdl.handle.net/10932/00-0332-C1B2-A5E3-2A01-D.

German Web 2013 (deTenTen). <https://www.sketchengine.eu>

Intercorp. <https://intercorp.korpus.cz/?lang=en>

MERLIN Written Learner Corpus für Czech, German, Italian. <https://merlin-platform.eu>

ZuMult - Zugänge zu multimodalen Korpora gesprochener Sprache – Vernetzung und zielgruppenspezifische Ausdifferenzierung. <https://zumult.org>

Kurzbio: Carolina Flinz ist Juniorprofessorin für Germanistische Linguistik am Department Lingue e Letterature Straniere der Universität Mailand. Sie promovierte an der Universität Pisa mit einem Projekt im Bereich der deutsch-italienischen Lexikographie (2009) und von 2010 bis 2017 arbeitete sie dort als Lehrbeauftragte (Economia & Management; Tourismuswissenschaft). 2017 war sie Humboldt-Stipendiatin für Erfahrene Wissenschaftler an der Universität Mannheim (12 Monate) und 2019 war sie DAAD-Stipendiatin am Leibniz-Institut für Deutsche Sprache (2 Monate). Von 2018 bis 2021 hatte sie eine Projektstelle an der Universität Mailand. Ihre Forschungsschwerpunkte sind: Lexikographie, Lexik, Fachsprachen, Korpuslinguistik, DaF, Kontrastive Linguistik.

Anschrift:
Dr. habil. Carolina Flinz
Università degli Studi di Milano
Dipartimento di Lingue e Letterature Straniere
Piazza S. Alessandro, 1
20123 Milano (MI)
carolina.flinz@unimi.it