

## 8 Eckdaten zum empirischen Teil

Bevor die Analyse von Konstruktionen mit *weil*, *da* und *denn* im Sprachgebrauch beginnt, seien an dieser Stelle einige Eckdaten zum empirischen Teil angegeben. Zunächst werden die verfolgten Ziele ausformuliert, anschließend wird ein Überblick über die verwendeten Korpora und das Vorkommen der drei Konnektoren innerhalb derselben gegeben. Den Abschluss bildet ein Abriss über Aufbau und Vorgehensweise der Analyse. Dieses Kapitel dient somit der Orientierung und sammelt die zentralen Informationen zum empirischen Teil an einem Ort. Konkretere Fragestellungen, zu überprüfende Hypothesen sowie nähere Beschreibungen der Korpora und der verwendeten Methoden finden sich in den jeweiligen Analysekapiteln.

### 8.1 Analyseziele

Das übergeordnete Ziel dieser Arbeit und damit auch der nachfolgenden Analysen ist es, die KxG mit drei ausgewählten kausalen Konnektoren auf komplexere, auf den ersten Blick ‚kerngrammatische‘ Phänomene auszuweiten und sie auf ihren Mehrwert für solche Phänomene hin abzuklopfen. Das symbiotische Ineinandergreifen von relativ stabilen Konstruktionen am Grunde des konstruktionalen ‚Flussbettes‘ und temporär, ‚angeschwemmten‘ *Frames* in Form bedeutungsseitiger *Construals* steht dabei im Vordergrund. Auf Basis bisheriger Studien und theoretischer Annahmen wurde dafür bereits ein Modell vorgestellt, das in der Analyse erprobt werden soll. Im Vordergrund steht also die Frage, inwiefern das Modell den Sprachgebrauch inklusive seiner Unterschiede in verschiedenen *Communities* erklären kann und in welchen Fällen es einer Ergänzung oder Überarbeitung bedarf.

Die skizzierten Konstruktionen und Konstellationen sind zum jetzigen Zeitpunkt somit als vorläufig zu verstehen. Zwar wurden bereits einige stützende Befunde aus der bisherigen Forschung angeführt, jedoch wurden diese nicht mit dem Ziel entwickelt, das hier vorgeschlagene Modell zu untersuchen. Vielmehr dienten sie der Beantwortung jeweils eigener Fragestellungen, womit eine vollständige Passung mit den hier zugrundegelegten Prämissen der KxG nicht immer gegeben ist. Beispielsweise wurden unterschiedliche Einbettungsvarianten bislang vorrangig unter dem Begriff der Fokus-Hintergrund-Gliederung untersucht, nicht aber bezüglich ihrer Topik-Kommentar-Struktur bzw. mit Blick auf das Konzept der *Aboutness*. Der Zusammenhang mit Verknüpfungsebenen wiederum wurde zwar syntaktisch und lexikalisch geprüft, jedoch noch nicht unter Einbeziehung der evozierten *Frames* beider Konnekte und deren Beziehung zum *Ground*, die ich für entscheidender halte.

Auch Korrelationen mit bestimmten Konstruktionentypen wurden noch nicht in der hier vorgeschlagenen Form erfasst, sondern etwa mit dem ‚Passiv‘ (z. B. Pit 2007: 60) nur ein Teilausschnitt aller nonagentiven Konstruktionen (vgl. u. a. Lasch 2016) oder mit ‚Pronomen der ersten Person‘ (z. B. Stukker & Sanders 2012: 177) nur ein Teilausschnitt von Konstruktionen mit Referenz auf die Sprecherin. Hier wird dagegen ein umgekehrtes Vorgehen angestrebt, das an der Bedeutungsseite ansetzt, anstatt vorab bestimmte Formen in den Blick zu nehmen.

## 8.2 Verwendete Korpora

Für die Analyse dienen drei zentrale Korpora als Grundlage, teilweise unterstützt durch einzelne Belege zu illustrativen Zwecken aus anderen Quellen, die jeweils gesondert angegeben werden. Die untersuchten Belege entstammen allesamt medial schriftlichen Gebrauchskontexten, da in derzeit vorliegenden gesprochen-sprachlichen Korpora keine hinreichende Treffermenge für jeden Konnektor hätte erzielt werden können, um quantitative Vergleiche zu erlauben.

Am umfassendsten wird dabei das *ZEIT*-Korpus untersucht: Diese Bezeichnung meint im Folgenden einen Ausschnitt des DWDS-Korpus *Die ZEIT (1946–2018)* (<https://www.dwds.de/d/korpora/zeit>), wobei nur Belege aus dem Zeitraum 2010 bis 2018 extrahiert wurden, um diachrone Entwicklungen ausklammern zu können. Der so gewählte Korpusausschnitt umfasst 238 Mio. Token und enthält online verfügbare zeitungssprachliche Texte, womit er eine schriftsprachliche Varietät abdeckt, die rezeptionsseitig potenziell der gesamten Diskursgemeinschaft des Deutschen zugänglich ist, also gewissermaßen dem ‚Hauptstrom‘ des deutschen Konstruktionsangehört.

Neben diesem Korpus ist außerdem das *Gingko* Gegenstand der Analyse (*GESCHRIEBENES INGENIEURWISSENSCHAFTLICHES KORPUS*, Version 1.0) (für einen Überblick vgl. Schirrmüller et al. 2021). In diesem Korpus, das 4,7 Mio. Token umfasst, sind automobiltechnische Fachartikel aus den Jahrgängen 2007 bis 2016 der Zeitschriften *Automobiltechnische Zeitschrift* und *Motortechnische Zeitschrift* versammelt, womit hier das Sprachverhalten einer sehr spezifischen, fachlich definierten *Community* erfasst wird. Mitglieder dieser *Community* müssen strenge Voraussetzungen erfüllen, sowohl um eigene Äußerungen produzieren zu dürfen (wofür ein abgeschlossenes Studium und die Mitarbeit in einer Forschungsgruppe nötig ist), als auch um die Äußerungen der anderen Mitglieder verstehen zu können (wofür diverse fachliche Konzepte und Vorgänge bekannt sein müssen).

Das dritte verwendete Korpus, hier als *Telegram*-Korpus bezeichnet, setzt sich aus 150.000 Nachrichten der öffentlich zugänglichen *Telegram-Supergroup Pokémon GO – Deutschland* (ca. 3.000 Mitglieder) zusammen, die zwischen Juli 2016 und Juli

2021 versendet wurden und über die *Sketchengine* maschinell annotiert wurden. Das Korpus umfasst 2,3 Mio. Token und zeigt eine dialogische, konzeptionell mündliche Varietät, die zugleich in einer thematisch klar abgegrenzten *Community* gebraucht wird, zu der aber im Vergleich zum *Gingko* ein niederschwelligerer Zugang besteht, insbesondere was die Produktion von Äußerungen betrifft. Tab. 7 gibt einen Überblick über die verwendeten Suchstrings und das Vorkommen der drei Konnektoren in den untersuchten Korpora.

**Tab. 7:** Verwendete Suchstrings und Vorkommen der Konnektoren in den untersuchten Korpora.

		<i>weil</i>	<i>da</i>	<i>denn</i>
ZEIT-Korpus (DWDS, 2010–2018)	Suchstring	weil with \$p = KOUS	da with \$p = KOUS	denn with \$p = KON
	Ausgabe	143747	41086	78317
	% Konnektor- Instanzen <sup>143</sup>	98,04	78,91	91,74
	Extrapolierte Gesamtzahl	140928	32419	71850
<i>Gingko</i> (Version 1.0, 2007–2016)	Suchstring	lemma = /weil/ & id = /sentence/ & #2_i_#1	lemma = /da/ & id = /sentence/ & #2_i_#1	lemma = /denn/ & id = /sentence/ & #2_i_#1 & #3.#2
	Ausgabe	843	2518	716
	% Konnektor- Instanzen	100	93,46	94,34
	Extrapolierte Gesamtzahl	843	2353	675
Telegram-Korpus ( <i>Pokémon Go –</i> <i>Deutschland</i> , 2016–2021, verwaltet über <i>Sketchengine</i> )	Suchstring	[lempos_lc = „(weil)-c“]	[lempos_lc = „(da)-c“]	[lempos_lc = „(denn)-c“]
	Ausgabe	4138	1721	280
	% Konnektor- Instanzen	100	80,16	80,65
	Extrapolierte Gesamtzahl	4138	1379	226

<sup>143</sup> Für diese Angabe wurde zunächst so lange annotiert, bis je 100 Instanzen des jeweiligen Konnektors markiert waren und im Anschluss die Anzahl der falschpositiven Treffer ermittelt.

Für das *ZEIT*-Korpus wurden jeweils 500 Belege pro Konnektor und Analyse schritt untersucht, für die anderen beiden Korpora jeweils 100. Am *ZEIT*-Korpus soll zunächst das theoretische Modell grundlegend erprobt werden, während für die anderen beiden Korpora weniger die exakten Verteilungen als vielmehr die grobe Richtung der Unterschiede zu den anderen Korpora von Interesse ist.

### 8.3 Kurzüberblick über Aufbau und Vorgehensweise

In den folgenden Kapiteln werden nacheinander mehrere ‚Tauchgänge‘ in unterschiedliche Tiefen und unterschiedliche Teile des ‚Sprachgebrauchsflusses‘ unternommen, wobei jeweils unterschiedliches Werkzeug nötig ist, um Aussagen über das zugrundeliegende Konstruktikon zu ermöglichen. Die grundsätzliche ‚Kartierung‘ erfolgt dabei zunächst mit dem *ZEIT*-Korpus anhand desjenigen Ausschnittes, der am ehesten den ‚Hauptstrom‘ des Deutschen abbildet, dem also rezeptionsseitig mehr oder weniger alle Mitglieder der Sprachgemeinschaft folgen können. Nacheinander werden dabei diejenigen Aspekte in den Blick genommen, die im theoretischen Teil als mögliche Schaltstellen zur Steuerung des Gebrauchs von *weil*, *da* und *denn* identifiziert wurden.

Im ersten Schritt (Kap. 9) werden die Stellungsvarianten in den Blick genommen, die nach dem entwickelten Modell nur indirekt mit den einzelnen Konnektoren und ihren *Construal*-Präferenzen verknüpft sind: Zunächst gilt es, mögliche Hinweise für einen Zusammenhang mit der *Aboutness*-Struktur herauszuarbeiten, die für alle Konnektoren gleichermaßen gelten und somit für eine grundsätzlich informationsstrukturelle Steuerung sprechen. Hierfür werden für die im *ZEIT*-Korpus vorkommenden Realisierungsvarianten von Konstruktionen mit *weil*, *da* und *denn* die Konnektlängen verglichen sowie das Vorkommen pronominaler und kontrastiver Komponenten im internen und externen Konnekt.

Das Herzstück der Analyse bilden jedoch die *Construal*-Konfigurationen, die im Anschluss ausführlich am *ZEIT*-Korpus untersucht werden. Als Methode hierfür wird die *Collostral*-Analyse vorgestellt, die in Anlehnung an die Kollexemanalyse nach Gries und Stefanowitsch (Stefanowitsch & Gries 2003; Gries & Stefanowitsch 2004a) das Vorkommen bestimmter Konnektoren und bestimmter *Construal*-Konfigurationen mit vs. ohne einander vergleicht (Kap. 10). Die verschiedenen Konfigurationen des *Subject of Conceptualization* (SoC), des *Object of Conceptualization* sowie der Beziehung des SoC zum Object werden für das externe und das interne Konnekt nacheinander beleuchtet, im Anschluss werden bestimmte Kombinationen aus externem und internem Konnekt untersucht (Kap. 11). Ziel ist, die Unterschiede zwischen den Konnektoren bezüglich ihrer *Construal*-Präferenzen

herauszuarbeiten und diese mit dem theoretischen Modell abzulegen. Varianten, die im Lauf der Annotation als mögliche eigenständige Konstruktionskandidaten hervorstechen, etwa weil ihre Form oder Bedeutung Auffälligkeiten zeigen, werden gesondert in den Blick genommen.

Nachdem dieser erste Praxistest des Modells anhand einer recht großräumigen Erfassung des linguistischen ‚Habitats‘ erfolgt ist, werden weitere Tauchgänge in etwas spezifischere Teilbereiche des deutschen Konstruktionskons unternommen: Dem *ZEIT*-Korpus werden nacheinander Analysen des ingenieurwissenschaftlichen Fachsprachenkorpus *Gingko* sowie eines eigens hierfür erstellten *Telegram*-Korpus gegenübergestellt (Kap. 13). Die Vorgehensweise der *Collostrual*-Analyse bleibt dabei bestehen, wobei insbesondere die Frage interessiert, ob sich hier jeweils dieselben bzw. ähnliche *Collostruals* herauskristallisieren wie zuvor. Sollte dies der Fall sein, spräche das für stabile Kopplungen aus Konnektor und *Construal*-Konfiguration, womit Konnektor-Präferenzen einzelner Varietäten systematische Effekte wären und nicht Teil der Konstruktionseinträge sein müssten. Andersherum sprächen deutliche Unterschiede in den *Collostruals* für varietäten-spezifische Ausprägungen der Konnektoren, die gelernt werden müssten, sprich konstruktional verankert wären. Stellungsvarianten werden in diesem ersten Korpusvergleich zwar erfasst, aber nicht in der vorigen Ausführlichkeit untersucht – lediglich für das *Telegram*-Korpus wird der Unterschied zwischen *weil*-Instanzen mit Verbletzstellung im internen Konnekt jenen mit syntaktisch autonomer Stellung gegenübergestellt, da hier in der bisherigen Forschung teilweise von unterschiedlichen Konnektoren ausgegangen wird.

Im letzten Schritt werden die in den vorigen Kapiteln annotierten Daten in ein gemeinsames Modell integriert. Während zuvor besonders die signifikanten Unterschiede zwischen den Konnektoren auf eng abgesteckten Einzelbereichen der *Construal* von Interesse waren, gilt es nun zu ermitteln, welchen Beitrag die untersuchten Variablen jeweils leisten, um die Entscheidung für oder gegen einen der drei Konnektoren vorherzusagen und damit mutmaßlich zu steuern. Zudem werden neben der *Construal* die mutmaßlich informationsstrukturell gesteuerten Variablen einbezogen, um die Größe ihres Einflusses zu bestimmen. Hierfür wird ein *Random Forest* genutzt (Kap. 14), der zur Vorhersage der Konnektoren die bedeutungsseitigen Variablen *SoC*, *Object* und Beziehung zwischen diesen für jedes Konnekt verwendet. Auf der Formseite kommen die Konnektlängen und deren Unterschied hinzu (externes minus internes Konnekt), sowie die Realisierungsvariante und das Korpus.

Es bleibt allerdings festzuhalten, dass die Analyse nur Hinweise auf die Tauglichkeit des Modells zur Beschreibung und Vorhersage von Sprachgebräuchen liefern kann, sie kann es nicht zweifelsfrei belegen. Da ‚das Konstruktikon‘ an und für sich nicht untersucht werden kann, sondern nur seine Auswirkungen auf das

Sprachverhalten, sind gewisse Unschärfen nicht zu vermeiden und die hier gezogenen Kategoriengrenzen überlappen einander notwendigerweise. Oberstes Ziel der Darstellung ist also nicht Exaktheit oder Vollständigkeit, sondern die größtmögliche Transparenz bezüglich der getroffenen Entscheidungen im Operationalisierungs-, Annotations- und Auswertungsprozess.