Originalveröffentlichung in: Lobin, Henning/Witt, Andreas/Wöllstein, Angelika (Hrsg.): Deutsch in Europa. Sprachpolitisch, grammatisch, methodisch. – Berlin [u.a.]: de Gruyter, 2021. S. 235-283. (Jahrbuch des Instituts für Deutsche Sprache 2020)
DOI: *https://doi.org/10.1515/9783110731514-013*

Anke Lüdeling/Hagen Hirschmann/Anna Shadrova/Shujun Wan (Berlin)

Tiefe Analyse von Lernerkorpora

Abstract: Die Sprache von Lerner/-innen einer Fremdsprache unterscheidet sich auf allen linguistischen Ebenen von der Sprache von Muttersprachler/-innen. Seit einigen Jahrzehnten werden Lernerkorpora gebaut, um Lernersprache quantitativ und qualitativ zu analysieren. Hier argumentieren wir anhand von drei Fallbeispielen (zu Modifikation, Koselektion und rhetorischen Strukturen) für eine linguistisch informierte, tiefe Phänomenmodellierung und Annotation sowie für eine auf das jeweilige Phänomen passende formale und quantitative Modellierung. Dabei diskutieren wir die Abwägung von tiefer, mehrschichtiger Analyse einerseits und notwendigen Datenmengen für bestimmte quantitative Verfahren andererseits und zeigen, dass mittelgroße Korpora (wie die meisten Lernerkorpora) interessante Erkenntnisse ermöglichen, die große, flacher annotierte Korpora so nicht erlauben würden.

1 Fragestellung und Hintergrund

Wenn wir uns die folgenden Beispiele, geschrieben von fortgeschrittenen Lernenden des Deutschen als Fremdsprache (aus den Korpora Falko und Kobalt-DaF, siehe Abschn. 1.3) anschauen, sehen wir zunächst oberflächennahe Fehler und Merkwürdigkeiten.[1] Dazu zählen beispielsweise orthografische Abweichungen (wie *ausgeshen* in (1a) oder *aüßern* in (1c)), Kasus- oder Argumentstrukturprobleme (wie *kann ich die Frauenbewegung für vieles dankbar sein* in (1a)), Flexionsprobleme (wie *müssten* in 1b) oder phraseologische Probleme (wie *Haushalt einhalten* in (1b)).

1 Das große und interessante Thema ‚Fehler' können wir hier nicht ausführlich besprechen. Wir verwenden den Begriff ‚Fehler' hier so, wie er in der Lernerkorpusdiskussion oft gebraucht wird. In anderen Artikeln haben wir uns ausführlich mit den theoretischen Hintergründen und der Modellierung von Abweichungen, Fehlern, Errors und Mistakes beschäftigt (für einen Überblick siehe Lüdeling/Hirschmann 2015).

https://doi.org/10.1515/9783110731514-013

Publikationsserver des Leibniz-Instituts für Deutsche Sprache
URN: http://nbn-resolving.de/urn:nbn:de:bsz:mh39-103337

(1a) Eins ist sicher; die Frauen aus der 68'er Generation haben einen wichtigen Kampf gekämpft. [...] Aus meiner Sicht kann ich die Frauenbewegung für vieles dankbar sein. Keiner kann ja sagen wie die Welt ohne die Frauenbewegung ausgeshen hätte, aber ich glaube bestimmt, dass die Frauenauffasung heute anders wäre, hätte es keine Bewegung gegeben. (Falko-Essay-Korpus L2, Text fkb024_2008_07; https://hu.berlin/bsp1a)

(1b) Man hielt sie eigentlich für nutzols. Sie hatten keinen richtigen Job. Sie müssten sich um ihre Kinder kümmern und die Haushalt einhalten. (Falko-Essay-Korpus L2, Text fk024_2006_07; https://hu.berlin/bsp1b)

(1c) Zu diesem Thema würde ich gerne meine Meinung aüßern und zugleich die Situation in China zeigen. Ich muss zuerst zugeben, dass das Leben heutiger Jugend sehr interessant aussieht. (Kobalt-DaF-Korpus L2, Text CMN_013_2011_03; https://hu.berlin/bsp1c)

Selbst wenn diese Auffälligkeiten nicht vorlägen, würde man merken, dass die Texte nicht von Muttersprachler/-innen verfasst sind. Woran liegt das? Was genau klingt anders? Und wie kann man solche Unterschiede finden und beschreiben? Die Ermittlung und Beschreibung ist ein erster Schritt, aber das genügt natürlich nicht – unterschiedliche Phänomene erfordern eine unterschiedliche formale und theoretische Modellierung. Dies wird seit einiger Zeit diskutiert und in dieses große Themengebiet ordnet sich unser Aufsatz ein. Wir möchten herausfinden, wie oberflächenferne abstrakte Muster in der Sprache von Lernenden einer Fremdsprache ermittelt und modelliert werden können. Wir werden dies anhand von drei exemplarischen Studien schlaglichtartig beleuchten. Bevor wir diese in Abschnitt 1.2 motivieren und in den Abschnitten 2–4 genauer betrachten, wollen wir kurz einige relevante Entwicklungen bei der Erstellung und Auswertung von Lernerkorpora skizzieren (Abschn. 1.1).

1.1 Was bisher geschah. Ein knapper Überblick zu Lernerkorpora und Lernerkorpusstudien

L2-Erwerb[2] wird seit mindestens dreißig Jahren korpusbasiert untersucht (siehe Granger 1998a für eine frühe Beschreibung und die Artikel in Granger/Gilquin/

2 In diesem Artikel gehen wir nicht auf den Erstspracherwerb ein, für den es natürlich auch umfangreiche Korpora gibt. Da, wo allgemeine Aussagen gemacht werden und eine Differenzierung nicht für die Ergebnisse einer bestimmten Studie nötig ist, werden wir den Begriff L2-Erwerb für alle Formen des Fremd- und Zweitspracherwerbs verwenden (gesteuert, ungesteuert, Ln).

Meunier (Hg.) 2015).[3] Dabei sind häufig genau die oben angesprochenen oberflächennahen Auffälligkeiten betrachtet worden. Daneben gibt es eine Reihe von Untersuchungen zu bestimmten Gruppen von Lexemen und Kollokationen, die in Korpora direkt als Oberflächenform auf der Wortebene gesucht werden können (für eine frühe Arbeit zu Kollokationen vgl. beispielsweise Granger 1998b; für lexikalische Unterschiede in einer Registerstudie vgl. z. B. Gilquin/Paquot 2008). Solche Untersuchungen sind enorm wertvoll und vermitteln Grundlegendes über Erwerbsverläufe und verschiedene andere Erwerbsaspekte. Viele zunächst plausibel klingende Vorstellungen, beispielsweise zu Transfer oder Interferenz, konnten durch Detailstudien verändert und präzisiert werden. Dies führt zu neuen oder veränderten Erwerbsmodellen (Granger 1996; Gilquin 2008; Osborne 2015). Aufbauend auf diese Ergebnisse gibt es etliche interessante Entwicklungen, von denen wir drei kurz skizzieren möchten.[4]

a) **Bessere Korpora:** Neuere Lernerkorpora unterscheiden sich in jeder Hinsicht von den ersten unsystematischen ‚Fehlersammlungen' ohne Kontext und auch von ersten Textsammlungen (vgl. z. B. Heringer 2001). Die Unterschiede betreffen sowohl das Korpusdesign (Welche Texte werden in ein Korpus aufgenommen?) als auch die Korpusarchitektur und -aufbereitung (Wie sind die Texte ausgezeichnet? Wie ist die Qualität der Annotationen gesichert?). Zum Korpusdesign lässt sich sagen, dass sich desto mehr relevante Parameter zeigen, je mehr Studien durchgeführt werden. Während man zunächst hauptsächlich auf die Parameter Medialität (geschriebene vs. gesprochene Sprache), die L1 der Lernenden und ihren Lernstand geachtet hat, weiß man inzwischen, dass – neben vielen weiteren Parametern – auch die Aufgabenstellung (siehe dazu bspw. das Stichwort *task effects*, Alexopoulou et al. 2017), andere (vorher gelernte) Ln (also L1, L2, L3 usw.), Umgebungs-

3 Wenig überraschend beschäftigen sich die meisten vorliegenden Lernerkorpusstudien mit dem Englischen als Fremdsprache. In den vergangenen Jahrzehnten sind immer mehr und bessere Korpora auch zu anderen Sprachen gesammelt worden. Neben Korpora geschriebener Sprache werden auch mehr ‚stumme' gesprochene Korpora ohne Audiosignal und multimodale gesprochene Korpora gesammelt (Ballier/Martin 2015; Sauer/Lüdeling 2016).

Die Learner Corpus Association (www.learnercorpusassociation.org, Stand: 26.8.2020) sammelt und bündelt die Aktivitäten rund um Lernerkorpusstudien (dort findet sich eine ausführliche Bibliografie) und veranstaltet die alle zwei Jahre stattfindenden Learner Corpus Research Conferences. Eine ständig aktualisierte Liste von Lernerkorpora findet sich unter https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html (Stand: 24.9.2020).

4 Auf Formate und Werkzeuge und auch auf die so relevanten Aspekte wie Zugänglichkeit und Transparenz können wir hier aus Platzgründen nicht eingehen.

parameter und sogar die Persönlichkeitsstruktur der Lernenden zu Unterschieden in der sprachlichen Realisierung beitragen. Alle diese Faktoren interagieren zusätzlich auf komplexe und fluide Weise miteinander (Dörnyei 2017).

Es gibt neben großen ‚allgemeinen' Korpora wie dem sehr einflussreichen ICLE[5] mehr und mehr auf bestimmte Fragestellungen gezielt und systematisch erhobene Sammlungen. Neuere Korpora enthalten im Allgemeinen erheblich mehr Metadaten als frühere und können entsprechend spezifischer ausgewertet werden. Auch in der Aufbereitung gibt es erhebliche Fortschritte. Zunächst wurden relativ allgemeine Fehlertagsets verwendet, die Oberflächenfehler klassifizierten. Erst langsam begann eine Diskussion um die theoretische Verortung der Phänomene und den nötigen Grad der Interpretation der Sprachdaten. Heute ist die Vorstellung von der Klassifizierung und allgemeinen Interpretation von Fehlern erheblich differenzierter. Dies liegt zum einen daran, dass inzwischen verstanden wurde, dass jede Annotation von Fehlern (wie jede andere Annotation) eine Interpretation der Daten ist und auf einer impliziten oder expliziten Vorstellung einer ‚korrekten' Äußerung beruht (zur Gegenüberstellung verschiedener Fehlerklassifikationen und ihrer korpusbasierten Anwendung vgl. Lüdeling/Hirschmann 2015). Zum anderen wird heute generell viel mehr auf die Transparenz und Qualität von Annotationsprozessen geachtet. Hierdurch werden auch immer öfter Annotationsrichtlinien und Evaluationsmetriken für die Korrektheit oder Übereinstimmung von Annotationen veröffentlicht (vgl. Larsson/Paquot/Plonsky ersch. demn.).

b) **Bessere quantitative Modelle:** Verständlicherweise werden in frühen Arbeiten überwiegend absolute oder relative Frequenzen einzelner Oberflächenstrukturen (z. B. von Wörtern oder Fehlerkategorien) ermittelt. Schon früh bestand jedoch der Anspruch, sich zwecks methodologischer Reifung auch einer stärker mathematischen Beschreibung zu nähern. Dies wurde überwiegend in Form von quantitativen Modellen zur Beschreibung von Wachstumsprozessen sowie im Framework frequentistischer Statistik umgesetzt. Häufig werden hier, wie auch in anderen textnahen Fächern üblich (z. B. den Sozialwissenschaften), relative Frequenzen als Realisierungen stochastischer Systeme beschrieben. Das Ziel besteht dabei darin, einerseits komplex verwobene multifaktorielle und graduelle Phänomene quantitativ zu erfassen, andererseits eine möglichst hohe Reliabilität der Ergebnisse durch eine Abgrenzung zu Zufallsergebnissen zu erreichen (Signifikanzberechnungen).[6]

5 https://uclouvain.be/en/research-institutes/ilc/cecl/icle.html (Stand: 26.8.2020).
6 Auf die Problematik dieses Konzepts können wir hier aus Platzgründen nicht eingehen. Eine gute Zusammenfassung bietet Koplenig (2019).

Relevant ist darüber hinaus aber, dass mathematische Modelle auch eine präzise Formulierung des zu untersuchenden linguistischen Phänomens erfordert – ungefähre oder konzeptuell vage Gedanken müssen für den Einsatz statistischer Modelle geschärft und entwickelt werden. Somit ist die Entwicklung quantitativer Modelle, die heute vor allem die Integration von gemischten Modellen umfasst (mixed-effects modeling, generative additive modeling), auch und vor allem eine linguistische Aufgabe.

Betrachtet man diesen Aspekt jedoch im Kontext des vorherigen Punkts a), wird schnell deutlich, dass dem enge Grenzen gesetzt sind: Ist der Vergleich möglichst homogener Gruppen erwünscht, die von sehr vielen Faktoren beeinflusst sind, und/oder bedarf es zugleich einer tiefen linguistischen Annotation, so wird die Menge verfügbarer Daten zunächst kleiner. Der kreuzweise Gruppenvergleich, z. B. von geschriebener vs. gesprochener Sprache in L1 vs. L2 (4er-Matrix) erfordert an sich schon eine höhere Datenmenge, um befriedigende Effektstärken, also überzeugende statistische Ergebnisse, zu erreichen. Dort, wo Daten eher idiosynkratisch und in ihrer Konkretheit schlecht oder gar nicht replizierbar sind (z. B. historische Daten oder opportunistisch gesammelte Daten, beispielsweise aus Sprachkurskontexten), haften der statistischen Betrachtung darüber hinaus konzeptuelle Probleme an, sofern Wahrscheinlichkeit (in typisch frequentistischer Auslegung) als idealisierte relative Frequenz über theoretisch unendlich wiederholte, real in dieser Art Daten jedoch nicht wiederholbare Zufallsexperimente verstanden wird.

Mit anderen Worten: Wenn die Daten so sind, wie wir sie linguistisch haben wollen – nach zahlreichen Kategorien fein differenziert und linguistisch tief annotiert – können wir mit diesen Daten nicht immer Statistiken rechnen. Wenn die Daten ausreichend groß für überzeugende Effektstärken sind, sind sie auf der linguistischen Ebene häufig nicht ausreichend differenziert. Mit dieser Erkenntnis besteht nunmehr ein Bedarf an der Entwicklung quantitativer Modelle auch für komplexe, tief annotierte und fein differenzierte linguistische Phänomene in kleineren Datenmengen.

(c) **Vernetzung von Methoden:** Ein dritter wesentlicher Aspekt, der eng mit (a) und (b) zusammenhängt, ist die Erkenntnis, dass jede Korpusstudie (genau wie andere Studien) Grenzen hat und jedes Korpusdesign sowie jede qualitative oder quantitative Untersuchung nur einen Teilaspekt eines relevanten Forschungsziels untersuchen kann.

Ein entscheidender Teil hiervon ist unspezifisch für die Linguistik und liegt im sogenannten Regress des Experimentleiters begründet (*experimenter's regress*; Franklin 1999; Collins 2016; Feest 2016): Unabhängig davon, wie oft ein Ergebnis mit Hilfe einer bestehenden Methode auch auf neuen Daten repliziert wird, bleibt

es unmöglich, Ergebnis und Methode zu trennen. Somit kann ein Ergebnis nur als gesichert angenommen werden, wenn es auch mit anderen Methoden gezeigt werden kann, da sonst eine Rückführung auf (mathematische, technologische, datentypspezifische, ...) Artefakte nicht ausgeschlossen werden kann. Die besondere Herausforderung in der Linguistik liegt jedoch erstens in der Komplexität der Unterscheidung von natürlichen vs. quasi-experimentellen vs. experimentellen Daten, zweitens in der Zahl der wirkenden Einflussfaktoren, drittens in der generellen Unwiederholbarkeit mit denselben Sprecherinnen und Sprechern (weil die Messung selbst Einfluss auf das innere Sprachsystem der Teilnehmer/-innen haben kann) sowie an weiteren einschränkenden Faktoren. Diese lassen sich nicht mit jeder Methode gleich gut isolieren: Beispielsweise können bestimmte Umgebungsfaktoren in Experimenten besser kontrolliert werden, sie unterliegen allerdings auch einem Trade-Off mit der Natürlichkeit der elizitierten Daten. Somit ergibt sich in der linguistischen Empirie ein besonderer Bedarf der externen Triangulation verschiedener Methoden, um Ergebnisse nachhaltig zu sichern. Jede Methode wirkt dabei wie ein eigenes Prisma, durch das die sprachliche Realität unterschiedlich gebrochen wird, wobei sich das Signal nicht verändert, die Einblicke aber sehr unterschiedlicher Art sein können. Daher werden heute verschiedene Untersuchungen kombiniert (bspw. verschiedene Metriken auf denselben Daten, Untersuchungen an unterschiedlichen Korpora oder Korpusstudien mit psycholinguistischen Studien und anderen Methoden; vgl. Gilquin 2008; Littré 2015 unter vielen anderen).

1.2 Formulierung von Leithypothesen und Motivation der drei Fallstudien

In diesem Artikel möchten wir an die im vorigen Abschnitt besprochenen Entwicklungen und Diskussionen anschließen und mit drei Fallstudien auf grundlegende zukunftsweisende Probleme eingehen, die durch die Hypothesen 1.2.1–1.2.3 (siehe unten) benannt werden. Die Fallstudien betreffen zunächst einmal ganz unterschiedliche linguistische Teilbereiche: In Kapitel 2 wird eine Analyse des Gebrauchs von verschiedenen Modifikatorklassen diskutiert (vor allem Hirschmann 2015). Kapitel 3 stellt eine neue Analysemöglichkeit für die Quantifizierung von Koselektionsbeschränkungen vor, wie sie eine Rolle in der Konstruktionsgrammatik und Phraseologie spielen (Shadrova 2020). Kapitel 4 behandelt den Gebrauch von rhetorischen Strukturen in argumentativen Texten (Wan i. Vorb.). Alle drei Phänomene sind von reinen Oberflächenformen abstrahiert: Man kann sie nicht finden, indem man nach bestimmten Wörtern sucht, sondern sie bedürfen einer weitergehenden linguistischen Kategorisierung und Modellierung. Alle

drei Phänomene sind außerdem linguistisch komplex und vereinen Aspekte aus verschiedenen – etwa syntaktischen und semantischen, lexikosyntaktischen, pragmatischen, oder textlinguistischen – Analysen. Zuletzt unterliegen alle drei Phänomene Lernprozessen im Spracherwerb, die über den Fremdsprachenunterricht hinausgehen. Das liegt teilweise an den bereits genannten Punkten: Ein Phänomen, das schon linguistisch schwer beschreibbar ist, lässt sich nicht ohne Weiteres didaktisch umsetzen. Teilweise liegt die Schwierigkeit der behandelten Phänomene für den gesteuerten Erwerb auch daran, dass sie neben sprachlichen (grammatischen) Aspekten auch auf vielfältige, oft subtile und teilweise idiosynkratische Weise mit kulturellen Aspekten der Zielsprache verwoben sind. In diesem Beitrag geht es uns jedoch darum, wie wir zunächst einmal in der Linguistik ein besseres Bild von den ausgewählten Phänomenen erlangen können.

Im Zuge der Phänomenbetrachtungen stützen wir uns auf die folgenden drei methodischen Hypothesen, auf die wir im weiteren Verlauf immer wieder zurückkommen werden (siehe 1.2.1-1.2.3).

1.2.1 Abstraktion und Vernetzung

Viele linguistische Analysen und Hypothesen (wie die hier behandelten zum Spracherwerb) lassen sich nicht auf Oberflächenphänomene reduzieren; es braucht ein vernünftiges Maß an Abstraktion. Dies liegt daran, dass die meisten linguistischen Phänomene – wir zeigen dies an den Phänomenen Modifikation, Koselektion und rhetorischen Strukturen – nicht lexikalisch sind und allenfalls exemplarisch, aber nie umfassend durch einzelne Lexeme repräsentiert werden können. Außerdem unterscheiden sich L2-Varietäten von L1-Varietäten auf allen linguistischen Ebenen. Um Wechselwirkungen zwischen verschiedenen Systemen (z. B. Syntax und Lexik) untersuchen zu können, müssen häufig verschiedene Betrachtungsebenen miteinander verknüpft werden. Aus diesen Gründen müssen die auszuwertenden Sprachdaten durch tiefe, differenzierte und deshalb immer auch manuelle Annotationen angereichert werden und es entstehen sehr komplexe Korpusdaten.

1.2.2 Modellierung

Abstrakte sprachliche Phänomene erfordern eine saubere, angemessene theoretische Modellierung. Eine konsequente Annotation erfordert die eindeutige Kategorisierung jedes Einzelfalls. Um das leisten zu können, ohne sich mit Vermischungen in Restkategorien zufrieden zu geben, müssen die berücksichtigten

Ebenen (z. B. Syntax, Semantik oder Pragmatik) klar definiert und in ihrer Wirkung auf das zu untersuchende Phänomen verstanden sein. Das bedeutet eine hochaufwändige theoretische Vorarbeit, bevor die eigentliche empirische Arbeit beginnt.

1.2.3 Mittelgroße Korpora

Tiefe, linguistisch informierte Analysen können nur auf kleinen bis mittelgroßen Korpora durchgeführt werden. Da linguistische Annotation ein Kategorisierungsprozess basierend auf häufig unscharfen, teilweise überlappenden Kategorien über potenziell ambige Strukturen ist, erfordert sie im Allgemeinen ein hohes Maß manueller oder teilmaschineller Bearbeitung und einen iterativen Verständnis- und Verständigungsprozess. Im Rahmen dieses Prozesses müssen alte Annotationen widerrufen und verbessert werden. Während große Korpora eine hohe Zahl und Vielfalt von Exemplaren (Beispielen) bieten, sind sie in aller Regel linguistisch nicht in dieser Weise beherrschbar. Sie bleiben deshalb zwangsläufig hinter den theoretischen Ansprüchen der Linguistik zurück. Kleine bis mittelgroße Korpora bedürfen im Gegenzug der Entwicklung besonderer quantitativer Methoden, die nicht von repräsentativer Größe ausgehen.

1.3 Lernerkorpora für unsere Fallstudien: Falko & Kobalt-DaF

Die in den Abschnitten 2–4 berichteten Studien basieren auf den Lernerkorpora Falko (https://hu-berlin.de/falko) und Kobalt-DaF (https://hu-berlin.de/kobalt-daf), die am Institut für deutsche Sprache und Linguistik der Humboldt-Universität zu Berlin zur Verfügung gestellt werden. Beide Korpora sind über das Suchinterface ANNIS (https://corpus-tools.org/annis/; Krause 2019) u. a. in einer online durchsuchbaren Instanz (https://hu-berlin.de/annis-falko) zugänglich.

Falko (**F**ehler**a**nnotiertes **L**erner**ko**rpus) besteht in seiner Kernversion aus schriftlichen Erörterungstexten, geschrieben von fortgeschrittenen Lernenden des Deutschen als Fremdsprache mit diversen Muttersprachen und Sprachbiografien. Das Korpus enthält neben den primären Textdaten auf jeweils getrennten Beschreibungsebenen zahlreiche Annotationen, die sowohl die grammatischen Äußerungsteile der Lernenden als auch die ungrammatischen oder stilistisch markierten Teile beschreiben. Auf diese Weise können systematisch bestimmte morphosyntaktische oder spracherwerbsspezifische Abweichungs- bzw. Fehlerkategorien gefunden werden. Zu Falko gehört ein L1-Vergleichskorpus von Muttersprachler/-innen des Deutschen, die dieselbe Schreibaufgabe bearbeitet haben.

Die Architektur und die methodischen Entscheidungen bei der Datenaufbereitung sind in Lüdeling et al. (2008) und Reznicek et al. (2010) detailliert beschrieben.

Die in den vorgestellten Analysen ausgewerteten Korpusdaten sind komplex annotiert und enthalten neben den primär von den Lernenden geschriebenen Wortformen grammatische Annotationen derselben sowie eine „Korrekturebene", die sich Zielhypothese nennt. Auf dieser werden für ungrammatische Ausdrücke grammatische Pendants formuliert. Die Zielhypothesen werden als Paralleltext behandelt und genau wie die originalen Lerneräußerungen grammatisch annotiert. Tokenbasierte Annotationen sind eine Lemma- und eine Wortartenebene. Unterschiede zwischen Lernertext und Zielhypothese lassen sich systematisch finden, womit sich auch bestimmte grammatische Abweichungen (Fehlerkategorien) gezielt finden lassen. Die Zielhypothesenebene ist zudem mittels Dependenzannotationen syntaktisch geparst, indem für jede Wortform die jeweils hierarchisch nächsthöhere Wortform spezifiziert und die Funktionsbeziehung zwischen beiden Formen angegeben wird (Foth 2006).

Das stark an Falko angelehnte Lernerkorpus Kobalt-DaF (Zinsmeister et al. 2012) folgt im Grunde demselben Aufbau. Der entscheidende Unterschied ist, dass die Lernerkohorten in Kobalt-DaF homogen organisiert sind: Das Korpus beinhaltet Texte chinesischer, schwedischer und belarussischer Deutschlernender zu vergleichbaren Anteilen und ebenso ein L1-Vergleichskorpus. Die Aufgabenstellung ist in Kobalt-DaF zusätzlich einheitlich gehalten; es handelt sich um argumentative Essays zur Frage „Geht es der Jugend heute besser als früheren Generationen?". Die Kobalt-Daten wurden ursprünglich mit dem Ziel erhoben, eine möglichst homogene Lerner- und Lernerinnengruppe abzubilden, weshalb nur Lerner/-innen mit einem onDaF-Testergebnis zwischen 115 und 129 Punkten einbezogen wurden, was etwa einem CEFR-Niveau von B2.2 entsprechen soll (zum onDaF-Test, heute onSET siehe Eckes 2010). Darüber hinaus wurden jedoch zusätzlich Daten ober- und unterhalb dieses Limits erhoben. Insgesamt liegen somit 151 Texte von belarussischen und chinesischen Lerner/-innen vor (89 BEL, 62 CH), sowie zusätzlich 20 L1-Vergleichstexte. Die Texte, die nicht für das Basiskorpus vorgesehen waren, stehen einschließlich der Zielhypothesen und diverser Annotationen in einem Zenodo-Repositorium zur Verfügung (DOI: 10.5281/zenodo.3584091).

Im Rahmen der in den Kapiteln 2–4 vorgestellten Studien wurden weitere, hier noch nicht genannte Merkmale (Annotationen) der Korpora genutzt oder zusätzlich geschaffen. Sie finden in den jeweiligen Kapiteln Erwähnung.

2 Modifikation

2.1 Zielsetzung

Traditionell wurde Zweit- oder Fremdspracherwerb unter der Maßgabe betrachtet, dass die morphosyntaktischen Regeln einer Sprache erfasst und bei der Sprachproduktion berücksichtigt werden. Vor diesem Hintergrund gilt das Deutsche als Fremd- oder Zweitsprache als relativ komplex, was die nominale und verbale Flexion und die Wortstellung angeht. Es ist naheliegend, dass sich Spracherwerbsmodelle und empirische Spracherwerbsbeobachtungen an den grammatisch omnipräsenten und obligaten Merkmalen und somit an der rektionsbedingten Flexion und der Topologie orientieren. So hat die Forschungsgemeinschaft ein relativ ausdifferenziertes Bild davon, welche Meilensteine bei der Beugung der Lexeme und ihrer Positionierung im Satz überwunden werden müssen, welche von der Zielsprache abweichenden Merkmale inwiefern sprachstandspezifisch sind usw. Natürlich ist dabei die Diskussion um den Status bestimmter grammatischer Merkmale häufig nicht abgeschlossen. Ein Beispiel ist die Positionierung der (finiten) Verben: Umstritten ist, ob es beim Erwerb des Deutschen als Fremd- oder Zweitsprache eine allgemeine, L1-unabhängige Erwerbsreihenfolge gibt oder je nach L1 der Lernenden oder zuvor erworbenen L2 Transfereffekte berücksichtigt werden müssen (vgl. hierzu u. a. Haberzettl 2005; Winkler 2011).

Schaut man auf grammatisch fakultative oder flexible Einheiten der Sprache, heißt dies keinesfalls, dass es sich hierbei um leichter zu beschreibende und erwerbende Strukturen handelt als diejenigen Strukturen, die in jedem Satz zu erwarten sind und als der grammatische Kern des Deutschen aufgefasst werden. Im Gegenteil: Spätestens seit den 1980er Jahren ist bekannt, dass es sich bei den Modalpartikeln des Deutschen, die einst als überflüssige „Flickwörter“ (Reiners 2004, S. 241) abgetan wurden, um ein Lernproblem des Deutschen handelt. In bestimmten Äußerungskontexten bzw. Registern bieten Modalpartikeln eine effektive Möglichkeit, Sachverhalte in verschiedene Modalitätskontexte zu setzen. Die von Muttersprachlerinnen und Muttersprachlern mitunter häufig verwendeten Ausdrucksmittel sind jedoch semantisch schwer zu fassen, ihre Platzierung bzw. Sequenzierung im Satz ist auch nicht trivial, und so ist bekannt, dass Modalpartikeln ein Lernproblem des Deutschen sind (vgl. Zimmermann 1981; Jiang 1994; Möllering 2004; Vyatkina 2007). Dies wirkt sich in erster Linie darin aus, dass die einzelnen Formen (*ja*, *doch*, *halt* usw.) von Lernenden im Vergleich mit Muttersprachlerinnen und Muttersprachlern deutlich seltener (oder überhaupt nicht) verwendet werden. Man spricht von einem relativen Mindergebrauch (underuse), also einer quantitativen Abweichung im Gebrauch. Dass diese Abweichung im

L2-Gebrauch (gegenüber dem L1-Gebrauch) auch positiv ausfallen kann, zeigt z. B. Lorenz (1998), indem seine Auswertungen zeigen, dass fortgeschrittene Lernende des Englischen als Fremdsprache intensivierende Adjektive systematisch häufiger gebrauchen als Muttersprachlerinnen und Muttersprachler in vergleichbaren Produktionskontexten. Man spricht hier von einem relativen Übergebrauch (overuse). Andere Arbeiten haben gezeigt, dass es Lernenden einer Fremdsprache schwerfällt, registerspezifische Ausdrücke angemessen zu verwenden und der Ausdruck häufig konzeptionell zu mündlich ausfällt (Aijmer 2002; Gilquin/Paquot 2007, 2008). Im Vergleich zum Sprachgebrauch von Muttersprachlerinnen und Muttersprachlern kann man von qualitativ abweichender Verwendung oder auch stilistischem Fehlgebrauch (misuse) sprechen.

Diese Studien weisen stark darauf hin, dass im Bereich der Verwendung von Modifikatoren enorme erwerbsbedingte Effekte existieren, deren Untersuchung ebenso relevant erscheint wie der Erwerb rektionsbedingter, obligatorischer Strukturen. Die maßgeblich in Hirschmann et al. (2013) und Hirschmann (2015) vorgestellten Auswertungen des Falko-Korpus zeigen, dass man, um etwas über die Verwendung bestimmter syntaktisch-semantischer Funktionen aussagen zu können, möglichst von der Betrachtung konkreter Oberflächenformen abstrahieren muss. Bei den Analysen wird ersichtlich, dass sich wesentliche funktionale Kategorien (wie z. B. „Satzadverbial“, „Attribut“ oder „Fokusmodifikator“) schwerlich durch einzelne Formen abbilden lassen, sondern die zu analysierende Klasse allein durch eine Analyse sämtlicher Formen derselben funktionalen Kategorie befriedigend erfassen lässt. Hierbei entsteht immer ein variationistisches, gebrauchsbasiertes Bild.

2.2 Datenauswertung

Die im Folgenden dargestellten Auswertungsmethoden zielen also darauf ab, den Gebrauch von Modifikatoren bei Lernenden des Deutschen als Fremdsprache und deutschen Muttersprachler/-innen im Vergleich zu beleuchten und dabei möglichst von der Ebene einzelner Wortformen oder Grundformen zu abstrahieren, ohne aber die Möglichkeit außer Acht zu lassen, die lexikalische Ebene bei Bedarf mitzuberücksichtigen.

Zunächst bedarf es bei dem Vorhaben einer möglichst umfassenden Klassifikation von Modifikatoren. Hiermit befasst sich ein wesentlicher Teil aus Hirschmann (2015, S. 65–198), indem die traditionellen Unterscheidungsmerkmale der (modifizierenden) Wortarten und Satzglieder kritisch zusammengetragen und anhand vereinheitlichter Kriterien in Klassifikationen zusammengebracht werden. Dies erfolgt zunächst für die modifizierenden Wörter bzw. Wortarten (Modal-

wort, Modalpartikel, Adverb, Adjektiv, Grad- und Fokuspartikel) und anschließend für Phrasen, bestehend aus mehreren Wörtern (im Grunde sind hier sämtliche Phrasentypen von Relevanz). Als Resultat entsteht eine funktionale Klassifikation gemäß dem Modifikationsziel der Modifikatoren, die formunabhängig zwischen adpropositionalen, adverbialen, adnominalen, graduierenden und fokusassoziierten Modifikatoren unterscheidet und dabei berücksichtigt, dass nicht nur im verbalen, sondern auch im nominalen Bereich event-interne wie -externe Modifikatoren auftreten können und dass die meisten modifizierenden Ausdrücke des Deutschen polyfunktional sind und somit nicht per se einer bestimmten Klasse zugeordnet werden können.

Alleine wegen des letztgenannten Aspekts ist klar, dass man bei einer umfassenden Untersuchung von modifizierenden Ausdrücken nicht in erster Linie formbasiert vorgehen kann. So erfolgte die Annäherung an die Analyse von Modifikatoren über abstraktere Kategorien, und zwar Wortarten- und Dependenzkategorien, die in Falko verfügbar sind.

Nach der beschriebenen theoretischen Vorarbeit wäre wünschenswert gewesen, die besagten Modifikatorkategorien in den zu analysierenden Korpusdaten auszuweisen, um bei der Analyse eine maximale Aussagekraft der Kategorien zu erzielen. Es gehört zu den größten Limitationen der beschriebenen Forschungsarbeit, dass dieses Vorhaben im zur Verfügung stehenden Rahmen nicht realisiert werden konnte. In der Analyse des Falko-Korpus wurde versucht, dieses Defizit auszugleichen, indem die Aussagekraft der vorhandenen gröberen Kategorien im Korpus auf angemessen Weise verknüpft wurden und indem Daten auszugsweise nachannotiert wurden.

Die im Korpus zur Verfügung stehenden grammatischen Kategorien, die für die Analyse von Modifikatoren genutzt wurden, setzen sich zusammen aus Wortartenannotationen und syntaktischen Annotationen. Zuletzt wurden auch individuell verwendete Lemmata bei der Analyse berücksichtigt.

Die Wortartenannotationen in Falko bestehen aus den Klassen des STTS-Wortartentagsets.[7] Dieses System unterscheidet zwar adjektivische Modifikatoren nach pränominal-flektierten und prädikativ, adverbial und ad-adjektivisch unflektierten Formen und weist morphologisch erkennbaren Pronominaladverbien eine entsprechende Kategorie (PROAV) zu und ist somit in diesen Bereichen relativ differenziert. Für sämtliche unflektierbaren Adverbien, Modalpartikeln und anderweitige Partikeln sieht das Tagset jedoch nur eine „Restkategorie“ (ADV) vor. Es existieren verschiedene Vorschläge und Anwendungen zur Ausdifferenzierung dieser groben Klasse (z. B. Hirschmann 2011; Rehbein/Hirschmann 2014a, b).

7 www.sfs.uni-tuebingen.de/Elwis/stts/Wortlisten/WortFormen (Stand: 22.9.2020).

Mit Blick auf die Analyse gesprochener Sprache maßgeblich beim Aufbau des FOLK Korpus[8] wurde eine differenziertere Version des STTS-Tagsets geschaffen, die verschiedene Partikelklassen und Mündlichkeitsphänomene abbildet (vgl. Westpfahl et al. 2017). Diese Version passt jedoch nicht gut zum schriftlichen Register des analysierten Falko-Korpus und war zum Zeitpunkt der beschriebenen Analysen noch nicht verfügbar.

Bei der Analyse der Modifikatoren wurden zunächst die gröberen Klassen des herkömmlichen STTS ausgewertet. Anschließend wurde diese Analyse gekoppelt mit der Auswertung von syntaktischen Relationen zwischen verschiedenen Wortklassen. Die Abbildung 1 zeigt exemplarisch die Annotationen in Falko, die für die Auswertungen genutzt wurden: Zum einen stehen für die Auswertung die Wortartenkategorien des STTS (Zeile „ZH1pos“) zur Verfügung. Für das gezeigte Auswertungsszenario ist jedoch viel wichtiger, dass Falko Dependenzannotationen (oben rechts) zur Verfügung stellt (Grammatik nach Foth 2006, geparst mit dem MaltParser; Nivre/Hall/Nilsson 2006). Die exemplarische Suchanfrage (oben links) findet systematisch Fälle wie den gezeigten, nämlich Wörter, die Adverbiale (Kantenbezeichnung „ADV“) zu Verben (Wortartbezeichnung „V.*“) sind. Mittels dieser Anfrage werden sämtliche Wörter gefunden, die dieselbe Funktion wie *erfolgreich* im gezeigten Beispiel besitzen.

tok	um	erfolgreich	in	einem	Job	ageiren	zu	können
ZH1	um	erfolgreich	in	einem	Job	agieren	zu	können
ZH1pos	KOUI	ADJD	APPR	ART	NN	VVINF	PTKZU	VMINF
ZH1lemma	um	erfolgreich	in	ein	Job	agieren	zu	können

Abb. 1: Exemplarische Darstellung des Teilsatzes *um erfolgreich in einem Job agieren zu können* mit den für die Analyse von Modifikatoren relevanten Korpusannotationen

In dieser Vorgehensweise wurden die Gesamtvorkommen verschiedener Modifikatorklassen mit unterschiedlicher Granularität analysiert. Die Analysen wurden jeweils kontrastiv durchgeführt, indem relative Vorkommen im Falko-Essay-Ler-

8 http://agd.ids-mannheim.de/folk.shtml (Stand: 22.9.2020).

nerkorpus und dem entsprechenden Vergleichskorpus von Muttersprachler/-innen des Deutschen miteinander verglichen wurden. Das Ziel der Studien war es, Lernprobleme der Zielsprache Deutsch aufzudecken, die nicht ohnehin bekannt sind bzw. auf der Hand liegen. Hierbei ist auch der Hinweis zu beachten, dass das untersuchte Lernerkorpus Falko ausschließlich fortgeschrittene Lernende des Deutschen als Fremdsprache umfasst (dies sind grob eingestuft Lernende oberhalb des B2-Niveaus nach dem Gemeinsamen Europäischen Referenzrahmen (Council of Europe 2017)). Diese Lernenden können komplexe zusammenhängende argumentative Texte formulieren und erzeugen dabei eine relativ geringe Anzahl originär grammatischer Fehler. Somit sind robuste Effekte in der Sprachverwendung gegenüber muttersprachlichen Vergleichsgruppen umso interessanter und können auf subtile, von den Lernenden sowie Lehrenden und anderen Didaktiker/-innen unerkannte Probleme hinweisen. Wie wir sehen werden, birgt die Verwendung von Modifikatoren solche Effekte, die nach Erklärung verlangen.

2.3 Ergebnisse

Mit einem ersten groben Schritt wurden die in Falko verfügbaren STTS-Wortartenkategorien, die für bestimmte Modifikatoren stehen, systematisch verglichen. Dies betrifft die Klassen PTKNEG (Negationspartikel), PROAV (Pronominaladverb), ADJA (pränominal-flektiertes Adjektiv), ADJD (prädikatives, adverbiales oder ad-adjektivisches Adjektiv) und ADV (unflektierbare Adverbien, Modalpartikeln, Gradpartikeln und Fokuspartikeln). Interessanterweise sind alle diese Wortklassen in Falko-Essay L2 (den Lernerdaten) und Falko-Essay L1 (den muttersprachlichen Vergleichsdaten) absolut vergleichbar distribuiert. Lediglich die Verwendung der Klasse ADV weicht insofern ab, als dass sich ein signifikanter Mindergebrauch der DaF-Lernenden gegenüber den Muttersprachlern ergibt. Dies ist ein erster Hinweis darauf, dass bei der Verwendung adverbial und adsentential gebrauchter Modifikatoren und/oder verschiedenen Partikelklassen eine systematische Abweichung existiert, die unabhängig von der Verwendung einzelner Wörter ist. Wenn nämlich bestimmte „schwierige“ Wörter innerhalb einer bestimmten Wortklasse vermieden oder nicht erworben werden, so ist zu erwarten, dass bereits erworbene bzw. beherrschte Wörter umso häufiger verwendet werden, nicht aber, dass eine gesamte grammatische Klasse von Wörtern vermieden wird oder nicht erworben wurde. Genau hierauf deuten allerdings die Korpusdaten hin.

Die Tendenz von Lernenden des Deutschen als Fremdsprache, Modifikatoren minderzugebrauchen, die zunächst grob als adverbial oder adsentential zu verorten sind, ist in Hirschmann et al. (2013) beschrieben und wird in Hirschmann

(2015) aufgegriffen und genauer untersucht. Zunächst wird gezeigt, dass das abweichende Verhalten der Lernenden ein allgemeines Phänomen ist und nicht etwa durch einen bestimmten Transfereffekt bedingt ist: Da Falko Lernergruppen mit sehr verschiedenen Muttersprachen enthält, kann man das Korpus nach allen enthaltenen Gruppen separat auswerten. Hierbei wurde festgestellt, dass der gezeigte Effekt für alle Lernergruppen gilt.

Um weiter von einzelnen Wörtern zu abstrahieren und wirklich syntaktische Verwendungen zu untersuchen, wurden in Hirschmann (2015) verschiedene syntaktische Interpretationen (Parses wie in Abb. 1 gezeigt) der Falko-Daten ausgewertet. Hier bestätigt sich das Bild, dass sich der Mindergebrauch von Modifikatoren auf adverbiale und adsententiale Ausdrücke bezieht: In einem Vergleich der Falko-Essays (L2 und L1) mit den Zeitungsdaten des syntaktisch geparsten Tiger-Korpus[9] auf syntaktische Funktionen hin lassen sich sowohl Registerunterschiede als auch spracherwerbsspezifische Effekte ablesen. Gleichen sich die Falko-L2- und die Falko-L1-Gruppe und beide Gruppen unterscheiden sich gleichermaßen vom Zeitungskorpus, so deutet dies auf einen Registerunterschied (zwischen der stark formalisierten Zeitungssprache und den etwas umgangssprachlicheren Essays von jungen Schreibenden im Falko-Korpus) hin. Unterscheiden sich Die Falko-L1-Daten und die Daten des Tiger-Korpus gleichermaßen von den Falko-L2-Daten, so deutet dies auf einen spracherwerbsbedingten Unterschied hin. Hinsichtlich der syntaktischen Funktion „adverbiale oder adsententiale Modifikation“ gilt genau Letzteres, wohingegen andere Modifikationstypen (z. B. pränominale Adjektive und postnominale Präpositionalphrasen) registerspezifisch variieren.

Zum einen passen diese Ergebnisse, die allesamt darauf hinweisen, dass es einen sehr robusten Spracherwerbseffekt beim Gebrauch von Modifikatoren in der verbalen Domäne gibt, zu den bereits benannten Beiträgen, die die semantisch relativ abstrakten und unscheinbaren Modalpartikeln als Lernproblem im Deutschen ausmachen. Zum anderen bleiben aber die Gründe und somit die weiterreichenden Konsequenzen im Verborgenen: Ungeklärt ist, ob Lernende anstelle der nicht verwendeten Ausdrücke alternative Ausdrücke, die durchaus anderen grammatischen Klassen entsprechen können, verwenden, oder ob sie die durch die entsprechenden Modifikatoren ausgedrückte Bedeutung unspezifiziert lassen. Das letztgenannte Szenario wirft weitere Fragen auf, nämlich, ob die gegenüber den Muttersprachler/-innen unspezifischeren Äußerungen der Lernenden auf eine

9 www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/tiger/ (Stand: 22.9.2020).

geringer ausgeprägte Sprachkompetenz zurückgehen oder eher insofern kulturell verankert sind, als dass in verschiedenen Sprach- bzw. Diskussionskulturen ein unterschiedlicher Grad an Explizitheit oder auch Expressivität Usus ist.

In den weiteren Analysen sollte somit herausgearbeitet werden, welche syntaktischen Modifikatorklassen verantwortlich für die festgestellten Gebrauchsunterschiede bei modifizierenden Wortarten sind. Zum anderen wurde beleuchtet, ob der Mindergebrauch in irgendeiner Form kompensiert wird.

Um die erste Frage zu beantworten, wurden Modifikatorklassen als die Beziehung zwischen modifizierenden Wörtern (z. B. unflektierte Adjektive – vgl. die Wortkategorie „ADJD" von *erfolgreich* in Abb. 1), bestimmten Wortklassen von Dependenten (z. B. Vollverben – vgl. die Wortkategorie „VV" von *agieren* in Abb. 1) und modifizierenden Beziehungen zwischen Dependens und Dependenten (vgl. z. B. die Kantenkategorie „ADV" für „Adverbial" in Abb. 1) definiert. Auch die Position des Modifikators zum modifizierten Kopf wurde bei der Auswertung berücksichtigt. So konnten theoretisch fundierte und aussagekräftige Modifikationstypen wie die zwischen Fokuspartikel und fokusassoziierter Präpositionalphrase oder Nominalphrase, die zwischen postnominaler Präpositionalphrase und modifizierter Nominalphrase usw. unterschieden werden. Jeder Modifikationstyp wurde (quantitativ) auf seine relative Verwendungshäufigkeit und ggf. auch (qualitativ) auf die verwendeten Lexeme untersucht.

Als wichtigstes Ergebnis hat sich gezeigt, dass adpropositionale Einwortmodifikatoren, zu denen die bereits bekannten Modalpartikeln, aber auch epistemische (z. B. *offensichtlich*), evaluative (z. B. *leider*) und diskursverknüpfende (z. B. *infolgedessen*) Modalwörter (bzw. sog. Satzadverbien) gehören, verantwortlich für den allgemeinen Mindergebrauch modifizierender Ausdrücke bei den fortgeschrittenen Lernenden des Deutschen als Fremdsprache im Falko-Korpus sind.

Was die Frage betrifft, inwieweit die betroffenen Lernenden die mindergebrauchten Kategorien durch alternative Ausdrücke kompensieren, so lassen sich zweierlei Tendenzen ausmachen: Zum einen lässt sich die Tendenz der Lernenden belegen, adpropositionale Modifikatoren nicht adsentential (satzintern und satzbezüglich), sondern sentential (satzförmig und dem Bezugsatz übergeordnet) realisiert, weil adpropositionale Modifikatoren grundsätzlich extraponiert und als eigenständiger, übergeordneter Matrixsatz umformuliert werden können. Vergleiche hierzu die Beispielsätze (2a) und (2b) aus dem Falko-Korpus:

(2a) [...] aber es ist möglich, vermute ich, dass es globaliziert geworden ist . [...] (Falko Essay L2, Text fkb011_2008_07; https://hu.berlin/bsp2a)

(2b) Möglicherweise wird eines Tages sogar der gesamte Drogen- und Waffenhandel entlarvt [...] (Falko Essay L1, Text dew03_2007_09; https://hu.berlin/bsp2b)

(2a) stellt eine Lerneräußerung aus dem Falko-Essay-Korpus dar, bei der die Kernäußerung (sinngemäß: *Es ist globalisiert worden*) durch zwei vorangestellte Matrixsätze (*es ist möglich* und *vermute ich*) modifiziert wird. Der epistemische Gehalt dessen wird in der muttersprachlichen Äußerung (2b) durch das (adpropositionale) Modalwort *möglicherweise* ausgedrückt. Die durch die beiden Beispieläußerungen ersichtliche Tendenz lässt sich an den Falkodaten als systematisch nachweisen.

Diese Tendenz zur Variation adpropositionaler Ausdrücke und ihre Realisierung als sententiale Einheiten wird auch in Hirschmann (2013) anhand der Modalpartikel *ja* aufgezeigt: Die Deutschlernenden im Falko-Korpus verwenden anstatt der Modalpartikel *ja*, die in erster Linie die Proposition, in die sie eingebettet ist, als bekannt markiert, äquivalente satzförmige Ausdrücke wie *Es ist bekannt, dass...* Gleichzeitig finden sich in den Lernerdaten allgemein weniger Ausdrücke, die die Bekanntheitsbedeutung markieren.

Es ergibt sich also ein komplexeres Bild, das zunächst das bekannte Lernproblem der Modalpartikeln im Deutschen um adpropositionale Ausdrücke im Allgemeinen erweitert. Zusätzlich wird aber auch gezeigt, welche Strategien Lernende des Deutschen als Fremd- und Zweitsprache entwickeln können, um das zu Formulierende dennoch auszudrücken, und zwar nicht nur durch alternative Ausdrücke innerhalb derselben grammatischen Kategorie (das wären vor allem synonyme bzw. partiell synonyme Wörter innerhalb derselben Wortklasse), sondern durch Konstruktionen, die grammatisch vollkommen unterschiedlich zu beschreiben sind: Man beachte, dass adpropositionale Ausdrücke wie die Modalpartikel *ja* oder das Modalwort *bekanntlich* klare Modifikatoren sind, während der Ausdruck *Es ist bekannt, dass* der Kopf einer Komplementstruktur ist. Die beschriebenen Untersuchungen besitzen somit nicht nur spracherwerbstheoretische Implikationen, sondern auch grammatiktheoretische.

3 Koselektion

3.1 Zielsetzung

Das Phänomen der Koselektion beschreibt die gemeinsame Auswahl linguistischer Elemente oder Strukturen auf verschiedenen Ebenen. Beispiele koselegierter Strukturen sind Kollokationen (typischerweise gemeinsam gewählte Wörter, beispielsweise *starker Kaffee* statt *kräftiger Kaffee*) und Idiome (Kollokationen mit nichtkompositionellen Bedeutungsbestandteilen wie *auf den Putz hauen*), Chunks (syntaktisch und phonologisch nicht zwingend analysierte, holistisch gelernte

und abgerufene Wortgruppen), Kolligationen oder Kollostruktionen (lexikalisch spezifizierte Präferenzen in der syntaktischen Umgebung, z. B. Dativalternationen oder Präferenzen bei Genitiv- vs. Präpositionalgruppen zum Possessivausdruck), aber auch Konstruktionen, Verbargumentstrukturpräferenzen und registertypische Ausdrücke wie beispielsweise Funktionsverbgefüge (Koselektion von Textsorte und linguistischem Element). Bereits diese Auflistung spannt ein weites Feld linguistischer Kategorien und erwartbarer Verknüpfungen auf: Ursprünglich im Bereich der Korpuslinguistik hauptsächlich als lexikografisches Phänomen in der phraseologischen Forschung erfasst (Bahns 1993; Bartsch 2004; Sinclair 1991; Kilgariff/Tugwell 2001), beschäftigt sich nahezu die gesamte gebrauchsbasierte Linguistik in der einen oder anderen Weise (auch) mit Koselektion. Prominente Beispiele dafür sind lexikalisch teilspezifizierte Syntaxansätze wie die verschiedenen Varianten der Konstruktionsgrammatik (Goldberg 2006; Sag/Boas/Kay 2012; Croft 2001) und angelehnte Studien, etwa zu Kollostruktionsanalysen (Stefanowitsch/Gries 2003; Gries/Stefanowitsch 2004; Gries/Wulff 2005); die gebrauchsbasierte Spracherwerbsforschung in L1 und L2 einschließlich der Frage von (phraseologischer) Komplexität in L2 (Paquot 2018, 2019), und die Produktivitätsforschung, die sich mit der Erweiterbarkeit und Beschränkung von Koselektion beschäftigt (Lüdeling et al. 2017; Zeldes 2012, 2013).

Die Grundlage für dieses weitreichende Interesse liegt in der frühen gebrauchsbasierten Beobachtung, dass, trotz der formal unendlichen Generativität und potenziellen Originalität natürlicher Sprachen, die Menge der tatsächlich im Sprachgebrauch vorgefundenen Kombinationen vergleichsweise beschränkt zu sein scheint: Fast bekommt man den Eindruck, L1-Sprecher/-innen wählen aus immer derselben Menge vorgefertigter Äußerungen. Was davon abweicht, wird als unidiomatisch oder merkwürdig empfunden, gegenüber Lerner/-innen häufig mit dem Verweis, dass „man das so sagen *könnte*, aber nicht *würde*“. Diese Idee ist von Sinclair bereits 1991 als *idiomatisches Prinzip* (idiom principle) beschrieben worden, demgegenüber er das *Prinzip der freien Auswahl* (open choice principle) stellt. Es wurde allerdings bereits zuvor von Pawley/Syder (1983) als ein Rätsel im L1-Sprachgebrauch formuliert und auch im Kern in der gesamten Phraseologie immer wieder beschrieben. Theoretisch wird dabei ein sogenanntes phraseologisches Kontinuum postuliert, an dessen einem Ende idiomatisch, konventionell oder anderweitig festgelegte Elemente der Sprache stehen, während das andere Ende vollständig frei kombinierbare Elemente – etwa abstrakte grammatische Regeln – beheimatet (Granger 2005; Howarth 1998; Goldberg 1995 u. a.). In der L2-Erwerbsforschung wird Koselektion vorwiegend im Kontext von Chunks als frühe Lernhilfe und Herausforderung im Flexionserwerb einerseits, und in der Kollokationsforschung als Frage der phraseologischen Kompetenz fortgeschrittener Lerner/-innen andererseits diskutiert. Dabei dominiert die Beobachtung, dass

selbst weit fortgeschrittene Lerner/-innen große Schwierigkeiten beim Erwerb einer L1-ähnlichen Koselektion haben. Paquot (2019) beschreibt die phraseologische Komplexität sogar als bestimmendes Unterscheidungsmerkmal zwischen Lerner/-innen auf B2- vs. C1-Niveau nach den Maßstäben des Gemeinsamen Europäischen Referenzrahmens (CEFR; Council of Europe 2017).

Betrachtet man den Erwerb von Koselektion aus einer strukturellen Spracherwerbsperspektive, so wäre folgender Verlauf schlüssig: Lerner/-innen im frühen L2-Erwerb lernen häufig phrasenartige Konstrukte zum zügigen Aufbau einer basalen kommunikativen Kompetenz in der Fremdsprache, und auch didaktisch machen im gesteuerten Spracherwerb Aufgaben, in denen einzelne Wörter oder Konstituenten ausgetauscht werden, einen großen Teil aus. Weit fortgeschrittene Lerner/-innen haben viel Erfahrung mit der Zielsprache und insbesondere auch viel sprachlichen Input erhalten, und ggf. sogar bewusst und gezielt Kollokationen oder idiomatische Wendungen gelernt. Lerner/-innen auf mittleren Erwerbsstufen[10] hingegen sind mit einer starken Ausweitung der kommunikativen Situationen bei gleichzeitig höherer Komplexität konfrontiert, lernen eine Vielzahl Wörter und grammatischer Konstruktionen, haben aber noch wenig Erfahrung mit der Zielsprache. Da Koselektionsbeschränkungen bislang selbst linguistisch schwer in Regeln beschreibbar sind, es einen idiosynkratischen Anteil zu geben scheint, und der didaktische Fokus in der Regel auf anderen Aspekten des zielsprachlichen Systems liegt, liegt es nahe, dass diese Lerner/-innengruppe die größten Schwierigkeiten hat, zielsprachlichen Koselektionsbeschränkungen zu folgen. Mit anderen Worten: Wenn man genug weiß, um neu zu kombinieren, geht erstmal alles durcheinander. Es handelt sich somit strukturell um einen Prozess von Generalisierung, Differenzierung und Spezialisierung, der in einem angenommenen U-Kurven-Verlauf zu Lasten der Lerner/-innen auf mittleren Erwerbsstufen geht.

Während die generelle Beobachtung, dass Lerner/-innen weniger Koselektionsbeschränkungen beherrschen, leicht an der absoluten und relativen Zahl konkreter Kollokationsrealisierungen gemacht werden kann – dass Muttersprachler/-innen, mehr seltene und vielfältigere Kollokationen als Lerner/-innen verwenden, kann als gesichert angenommen werden (Laufer/Waldman 2011; Nesselhauf 2005; Paquot 2018 u. v. m.), ist die Quantifizierung und der Nachweis einer strukturel-

10 Der Begriff der Erwerbsstufe ist hier agnostisch verwendet und wird nur für eine grobe Einordnung benötigt. Ob es sich tatsächlich um einen kontinuierlichen Verlauf handelt, oder ob Erwerbsstufen sich durch spezifische grammatische, lexikalische usw. Eigenschaften auszeichnen, ist für die strukturelle Abstraktion zweitrangig. Gemeint ist eine relativ grobe Einteilung in CEFR-Stufen, etwa A/B/C oder B1/B2/C1, ohne den Anspruch einer exakten Verortung jedes Lerners oder jeder Lernerin.

len Verschiebung von Koselektionskompetenzen um einiges komplexer. Das liegt zum einen in der theoretischen Unschärfe durch die Dimensionsreduktion aller Formen festgelegter, konventioneller, oder idiomatischer Sprache in ein nur eindimensionales „phraseologisches Kontinuum“ begründet: Unterscheidet man nicht zwischen Elementen mit unterschiedlichen Eigenschaften, beispielsweise seltenen vs. häufigen Kollokationen, komplexen vs. weniger komplexen Chunks usw., so wird die Zählung erschwert: Syntaktisch motivierte häufige Wortkombinationen oder Chunks wie *das ist* erhalten beispielsweise denselben Status wie eng verknüpfte Idiome wie *ins Gras beißen*, obwohl sie aus linguistischer Sicht offensichtlich unterschiedliche Phänomene darstellen. Obwohl an letzterem wohl auch niemand zweifeln würde, stellt sich die Frage nach der Gewichtung in der quantitativen Modellierung jedoch für jedes einzelne Exemplar, was eine viel feingliedrigere Unterscheidung und ein weitaus expliziteres Modell erfordert als die bloße Zählung der Vorkommen. Darüber hinaus ist es denkbar und wahrscheinlich, dass die Phänomene in verschiedenen Dimensionen unterschiedlich ausgeprägt sind: Dass Lerner/-innen mittlerer Erwerbsphasen bereits Zugriff auf eine größere Sammlung häufiger Chunks haben, bestimmte Konventionen (z. B. Grußformeln) muttersprachenähnlich beherrschen, und dabei dennoch mit vielen Koselektionen nicht vertraut sind, die sich hinsichtlich ihrer Kompositionalität, Häufigkeit oder syntaktischen Flexibilität durchaus weitgehend unterscheiden können (beispielsweise syntaktisch feste Einheiten wie *Wer zuerst kommt, mahlt zuerst* oder syntaktisch flexible, aber registergebundene wie *Kritik üben* in *soll Kritik geübt worden sein*). Eine Sortierung aller möglichen Koselektionen oder konventionell gemeinsam verwendeten Einheiten auf einem monodimensionalen Kontinuum legt eine Modellierung des Erwerbs entlang desselben nahe (beispielsweise von fest zu frei, wie teilweise im L1-Erwerb angenommen; MacWhinney 2014; Tomasello 2009). Es ist aber nicht so, dass dieses Kontinuum tatsächlich auch in einer Richtung erschlossen wird, jedenfalls nicht im L2-Erwerb: Vielmehr werden zunächst Wörter und Konversationsformeln (fixe Einheiten) und grammatische Regeln (abstrakte Einheiten) gelernt, und ausgehend davon alle anderen Bereiche, diese aber eben nicht zu gleichen Anteilen. Es verbleibt als Aufgabe für die zukünftige Forschung, ein multidimensionales Modell auszuformulieren.

Für eine strukturelle Quantifizierung von Koselektion als Gesamtphänomen bereitet ein solches allein allerdings auch keine unproblematische Lösung: Würde man darin Dimensionen für Eigenschaften wie Konventionalität, Kompositionalität, Häufigkeit, morphosyntaktische Flexibilität und Idiosynkrasie anlegen, erhielte man bereits $2^5 = 32$ Kategorienkombinationen, in die alle gewählten Koselektionen zunächst einzusortieren wären – mehr noch, wenn man die Zugehörigkeit zusätzlich gradieren wollte (mehr oder weniger konventionell oder morphosyntaktisch flexibel). In dem Fall wäre man wieder dabei, Einzelfälle pro

Kategorie zu zählen und/oder müsste eine strukturelle Abstraktion vornehmen, die in dem Modell nicht naheliegend enthalten ist. Während ein solches Modell also hilfreich wäre für die Diskussion um phraseologische Komplexität bzw. die Beschreibung von Kompetenz in verschiedenen Teilbereichen von Koselektion, somit auch für die Erkundung möglicher Erwerbstrajektorien einschließlich didaktischer Konsequenzen und Lücken, ist es für die Quantifizierung des idiomatischen Prinzips eher von mittelbarem Nutzen. Es sei außerdem darauf verwiesen, dass selbst in einem thematisch sehr einheitlichen Korpus wie Kobalt-DaF die Zahl der identischen Koselektionen subkorpusübergreifend sehr gering ist (unter 10), während die allermeisten, auch zielsprachlich völlig akzeptablen und erwartbaren Koselektionen Hapax bzw. Dis Legomena sind, sprich nur ein- oder zweimal auftreten. Wegen der Spezifika von Lexemverteilungen in Korpora würde eine Erweiterung der Daten dieses Problem nicht lösen, sofern sie nicht mindestens eine, besser zwei Größenordnungen umfasste (sprich von 171 auf 1.700 oder 17.000 Texte): Es würden vorwiegend mehr neue Hapaxe auftreten, während der interessante Mittelbereich der Verteilung sich nur sehr schleppend auffüllen würde.

Hinzu kommt eine mathematische Problematik, die in Shadrova (2020, insb. Kap. 4) genauer ausgeführt wird und in der Potenzmenge von Wortkombinationen liegt: Im Allgemeinen werden Koselektionseigenschaften mit Hilfe wahrscheinlichkeitsbasierter Modelle quantifiziert, typischerweise in sogenannten *lexikalischen Assoziationsmaßen* (Gries 2013, 2019; Evert et al. 2017; Wiechmann 2008; Gregory et al. 1999). Obwohl zahlreich, basieren die meisten dieser Maße auf demselben Informationsgehalt von vier Zählungen im Korpus: Den absoluten Vorkommen der beiden Wörter einer Kollokation im Korpus mit- und ohne einander (das sind drei Felder der Matrix), und dem Vorkommen aller anderen Wörter im Korpus (das ist das vierte Feld). Über verschiedene Transformationen werden aus diesen Werten Berechnungsvarianten bedingter Wahrscheinlichkeiten abgeleitet, wobei die zugrundeliegende Frage stets lautet, *wie unwahrscheinlich eine Kombination gegenüber dem Zufall ist*. Kommt eine unwahrscheinliche Kombination – also die Kombination von ohnehin seltenen Wörtern, etwa *ad acta* – häufiger vor, als zufällig erwartbar, so schließt man daraus, es handle sich um eine vorher festgelegte Kombination, also nicht wirklich um unabhängige Wörter. Das gilt insbesondere dann, wenn die Wörter ohneeinander nicht oder nur sehr selten vorkommen, was mit sogenannten asymmetrischen Maßen wie z. B. DeltaP (Gries 2013) gemessen werden kann (z. B. hat *ad* auf *acta* wohl eine geringere Anziehung als *acta* auf *ad*, da *ad* auch in *ad infinitum*, *ad hoc*, *ad nauseam* usw. auftritt).

Problematisch ist diese Quantifizierung in mindestens dreierlei Hinsicht: Erstens gibt es berechtigte Zweifel am mathematischen Modell, da Wortfrequenzen wahrscheinlich nicht den Ansprüchen eines stochastischen Modells entsprechen (siehe Shadrova 2020, Kap. 4; Shadrova i. Dr.). Das soll hier nicht weiter

ausgeführt werden, es sei nur allgemein darauf verwiesen, dass Wahrscheinlichkeiten idealisierte Häufigkeiten in einem geschlossenen, pfadunabhängigen, unveränderlichen Zufallssystem sind, und das trifft auf Wörter schon deswegen nicht zu, weil man produktiv neue bilden kann.

Zweitens ist der Vergleich mit einer zufälligen Verteilung doppelt problematisch: Bereits in einem kleinen Teilkorpus von Kobalt-DaF ist die Potenzmenge der möglichen lexikalischen Kombinationen absurd hoch, nämlich um Größenordnungen höher als die geschätzte Zahl der Atome im Universum (Shadrova 2020, S. 105).[11] Gegen diese Potenzmenge ist jede Auswahl von Lexemen extrem unwahrscheinlich – das allein erlaubt aber keine Skalierung: Ein akzeptables Modell müsste plausibel erläutern, warum z. B. 1/10^14 ausreichend unwahrscheinlich ist, um als koselektionsbeschränkt zu gelten, 1/10^13 jedoch nicht. Somit erlaubt eine Herleitung aus der Unwahrscheinlichkeit des Vorkommens zweier Wörter gegenüber einer Zufallsverteilung keine plausiblen Rückschlüsse auf die *Unterschiede* in der Koselektionsbeschränktheit zweier Texte.

Dieser Problematik liegt ein Modellproblem zugrunde: Wo Koselektion als Kollokation, idiomatisches Prinzip, oder phonologisches/graphematisches Chunk verstanden wird, betrachten die bisher üblichen Modelle *Exemplare* oder *Kategorien von Exemplaren*: Es können nur Dinge gleicher Art gezählt werden. Allerdings liegt hier ein Zirkelschluss vor, denn das Gleichartige wird durch die (angenommen beschränkte) gemeinsame Auswahl bestimmt, während die (angenommene) Beschränktheit der gemeinsamen Auswahl aus der Gleichartigkeit abgeleitet wird. Somit können Kollokationen dann gezählt werden, wenn sie häufig sind, oder zuvor, beispielsweise in einem Wörterbuch, bereits als solche verzeichnet wurden, was letztlich die Koselektionsstruktur als Ähnlichkeitsabbildung auf lexikografische Arbeit definiert: Wie ähnlich ist die Verwendung bestimmter Wörter in ihrer Kombinatorik gegenüber dem, was die Lexikografie zuvor festgestellt hat? Es bedeutet auch, dass die Assoziationsmaße der einzelnen Lexeme einander gegenübergestellt werden müssen – *ad acta* verhält sich L1-ähnlich, aber *zur Verfügung stellen* nicht, *Ausbildung* kommt so häufig vor wie in L1, aber mit dem falschen Verb, usw. Dies entspricht einer Triangulation über tausende, in der lexikalischen Kombination über Milliarden Werte – folgt man Gries (2019), sogar über mehrere Maße – und erlaubt dennoch nur schwerlich, eine Gesamtaussage über die Koselektionsstruktur zu treffen. Was für Aussagen über einzelne Lexeme unausweichlich zu sein scheint (sofern man dem probabilistischen Modell folgen mag), ist jedoch für die Messung der strukturellen Eigenschaften nicht förderlich.

11 Unter Annahme extremer Beschränkungen immer noch mindestens 3.352 · 10^89 für die belegten Verb- und Akkusativlexeme in einem Subkorpus von 21 Texten vs. 10 ^78 bis 10 ^82 Atome.

Aus Sicht des Zweitspracherwerbs besteht jedoch Interesse über die konkreten Lexeme hinaus auch die strukturelle Entwicklung der Koselektionseigenschaften in L2-Texten messen: Beispielweise, um zu vergleichen, wie sehr sie dem idiomatischen Prinzip entsprechen, oder um zu erforschen, ob es einen strukturellen Umbau im Zweitspracherwerb gibt (analog zu Annahmen aus dem Erstspracherwerb, siehe z. B. Tomasello 2009; Wray 2002). Dafür benötigt man ein quantitatives Modell, das a) in der Quantifizierung auf den Vergleich mit der Zufallsziehung verzichten kann, und b) ohne Triangulation von abertausenden Maßen auskommt.

3.2 Datenauswertung

Eine solche Möglichkeit bietet die Modellierung als Netwerkgraph. Die Graphentheorie ist ein Teilgebiet der diskreten Mathematik, die sich mit den Elementen einer Menge (Knoten) und den Relationen dazwischen (Kanten) beschäftigt. Typische Beispiele für visualisierte Netzwerkgraphen sind U-Bahn-Fahrpläne oder Freundschafts- und Kontaktnetzwerke.[12] In der Linguistik dienen Graphen bislang hauptsächlich zur Visualisierung bestehender Analysen, etwa in Konstituenten- oder Silbenstrukturbäumen oder Taxonomien. Während es in der Frühphase der Linguistik einige graphtheoretisch fundierte Analysen gab (Goodman 1961; Jelinek/Bahl/Mercer 1975; Brainerd/Chang 1982), sind sie mittlerweile weitgehend aus der Kernlinguistik verschwunden, obwohl zahlreiche computerlinguistische Anwendungen intern auf Graphenstrukturen zurückgreifen. In anderen quantitativen Wissenschaften – insbesondere in den Natur-, jüngst aber auch zunehmend in den Sozialwissenschaften – nehmen Netzwerkanalysen und Graphmodellierungen jedoch einen zentralen Stellenwert ein (siehe auch Shadrova 2020, Kap. 5).

Für die Analyse von Koselektion lässt sich ein Korpus als Graph beschreiben, indem die gewünschten Lexeme, z. B. Verben und ihre Argumentkopflexeme, als Knoten und bestehende Dependenzverbindungen dazwischen als Kanten modelliert werden. Das erfordert natürlich zahlreiche Modellentscheidungen, beispielsweise was den Umgang mit Homographen oder Polysemie angeht, oder welche Lexeme einbezogen oder ausgelassen werden: Sollen beispielsweise für den Fall der Präpositionalobjekte die Präpositionen selbst enthalten sein oder nur ihre

12 Unter anderem auch Stammbäume, wobei diese zusätzliche begrenzende Eigenschaften haben, nämlich dass sie auf bestimmte Weise angeordnet (hierarchisch) sind, in der Regel azyklisch, und dass je ein Knoten einer Ebene mit genau zwei Knoten der darüberliegenden Ebene verbunden ist.

Nominalergänzungen – das hat Einfluss auf die Pfade zwischen Verben und Objekten, die im Korpus gar nicht zusammen auftreten, aber durch die Präposition verbunden werden. Die Entscheidung hat hierbei zum einen Einfluss auf die Abbildung des linguistischen Modells: Da jede Abstraktion einen Informationsverlust impliziert, ist eine Klärung der theoretischen Bedarfe der Forschungsfrage zunächst notwendig. Im Falle einer netzwerkanalytischen Betrachtung ist es jedoch auch ein möglicher Einfluss auf die Quantifizierung, der diese Entscheidungen im besonderen Maß erfordert: Graphen können mit sogenannten Graphmetriken vermessen und quantifiziert werden. Diese betreffen unterschiedliche Eigenschaften, etwa die einzelner Knoten (Zentralität, Zahl ein- und ausgehender Kanten), Verbindungsmuster (Durchmesser, kürzeste Wege), aber auch abstraktere strukturelle Eigenschaften, etwa Isomorphismen (strukturelle Gleichheit von Teilgraphen) oder innere Strukturiertheit (Modularität). Die Frage nach der Einbeziehung oder Auslassung bestimmter Verbindungen entspricht damit der Frage nach dem gewünschten, nämlich dem für die Messung angemessenen Abstraktionsgrad: An dieser Stelle sei auf unsere zweite Arbeitshypothese verwiesen: „Abstrakte sprachliche Phänomene erfordern eine saubere, angemessene, theoretische Modellierung“.

In der Analyse von Shadrova (2020) wird der Erwerb von Koselektionsbeschränkungen nicht (vorwiegend) als Eigenschaft einzelner Wörter betrachtet, sondern als strukturelle Eigenschaft des lexikalischen Gebrauchs in L1- bzw. L2-Texten. Soll heißen: Der Gebrauch aller Lexeme in realisierten Verbargumentstrukturen kann mehr oder weniger strukturiert, also nach Lexemen differenziert sein: Die untere Grenze der Koselektionsbeschränkung – nämlich keine – bildet ein Text, in dem alle Verben mit allen Argumentlexemen verbunden sind (*Die Bundeskanzlerin liest ein Buch. Die Bundeskanzlerin schreibt ein Buch. Die Bundeskanzlerin kauft ein Buch*). Die obere Grenze der Koselektionsbeschränkung bildet ein Text, in dem jedes Verb einzigartige Argumentlexeme wählt (*Die Bundeskanzlerin liest ein Buch. Der Autor schreibt einen Roman. Die Verlegerin kauft eine Zeitschrift*).

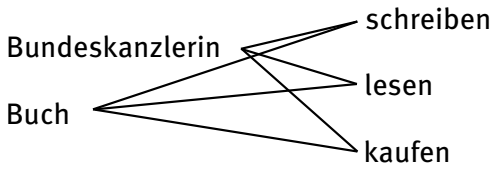

Abb. 2: Beispiel für einen vollständig verbundenen bipartiten Graphen von Dependenzbeziehungen

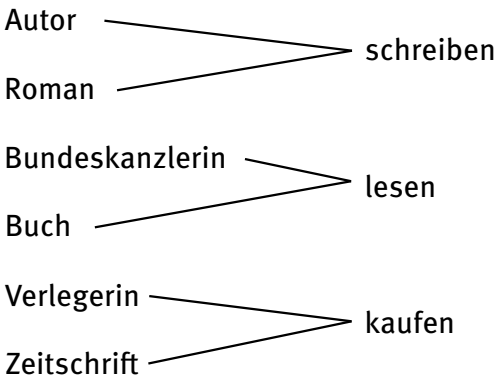

Abb. 3: Beispiel für einen in drei Subgraphen unterteilten Graphen von Dependenzbeziehungen

Diese beiden Graphen unterscheiden sich nicht nur in ihrem Inhalt, also darin, welche Knoten sie konkret enthalten, und in der Zahl ihrer Knoten, sondern auch auffällig in ihrer Struktur: Der erste Graph ist ein vollständiger bipartiter Graph, also ein Graph, der aus zwei Partitionen besteht (Verben und Substantive), wobei alle Elemente der einen Partition mit allen Elementen der anderen Partition verbunden sind. Der zweite Graph besteht aus drei unverbundenen Teilgraphen. Wichtig ist, dass sowohl im ersten, als auch im zweiten Beispiel alle Kombinationen semantisch möglich sind. Betrachtet man jedoch die Vorkommen in Korpora, so wird immer wieder festgestellt, dass eben nicht alle denkbaren Kombinationen vorkommen, und dass es vor allem auch große Unterschiede in der Frequenz des gemeinsamen Vorkommens auch bei semantisch und syntaktisch akzeptablen Strukturen gibt. Für das zweite Beispiel bedeutet das, dass zwar auch Bundeskanzlerinnen Bücher schreiben können und Zeitschriften kaufen, man aber selbst in einem sehr großen Korpus nicht alle Kombinationen dieser Wörter gleich häufig und einige möglicherweise gar nicht finden würde. Modelliert man einen Graphen dieser Art aus den 20 in Kobalt-DaF enthaltenen L1-Texten, entsteht ein visuell unüberschaubares Netzwerk wie in Abbildung 4.

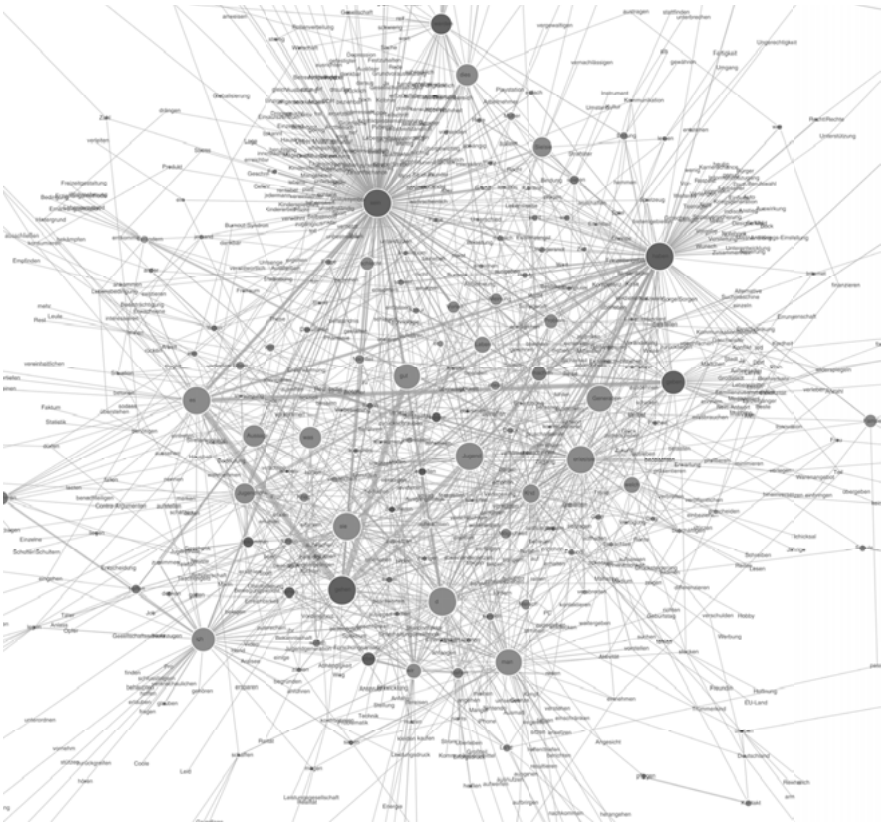

Abb. 4: Darstellung von Kobalt L1 als Netzwerkgraph. Helle Knoten entsprechen Verblexemen, dunkle Argumentkopflexemen. Darstellung dient nur zur Illustration des Prinzips, eine skalierbare Version ist über Zenodo (DOI: 10.5281/zenodo.3584091) verfügbar

Während die Knoteninhalte (Lexeme) kaum lesbar sind und die Struktur aus der bloßen Betrachtung, im Gegensatz zum obigen Minimalbeispiel, wenige Rückschlüsse auf die darunterliegende sprachliche Realität zu erlauben scheint, liegt das Potenzial dieser Modellierung in der Abstraktion, nämlich der Quantifizierung. Konkret erlaubt diese Art Graph die Berechnung der sogenannten *Modularität*, eines durch einen *Community-Detection-Algorithmus* ermittelten Maßes, das die innere Strukturiertheit eines Graphen beschreibt. Abbildung 5 zeigt das beispielhaft für drei kleine Graphen. Der hier verwendete Algorithmus wird als *Louvain-Modularität* bezeichnet (Blondel et al. 2008). Es gibt noch weitere Modularitätsmaße, die hier nicht weiter beachtet werden. Abbildung 6 und 7 zeigen im

Unterschied zu den vorherigen Beispielen drei gleich große Graphen (je 10 Knoten und 15 Kanten), die dennoch unterschiedlich stark intern strukturiert sind und entsprechend höhere bzw. niedrigere Modularitätswerte haben.

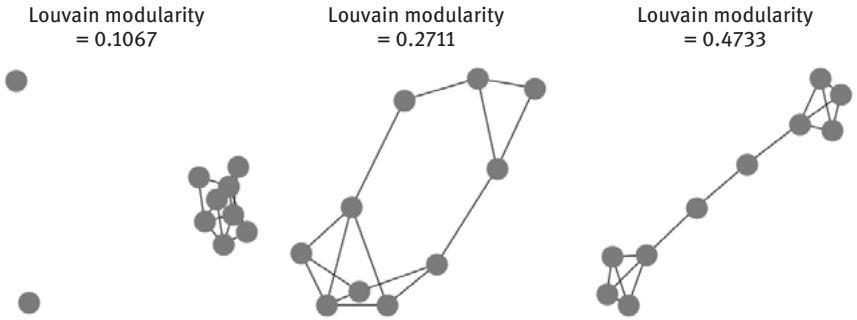

Abb. 5 **Abb. 6** **Abb. 7**

Abb. 5–7: Beispiele für Graphen mit verschiedenen Modularitätswerten

In der Studie ist anhand der Kobalt-Daten die Louvain-Modularität für Subkorpora von Lerner/-innen auf verschiedenen Sprachstufen, analog etwa zu CEFR A2-C1, berechnet worden. Die Subkorpora umfassten dabei zwischen 10 und 27 Texten. Die Hypothesen lauten, dass (1) Lerner/-innen weniger strukturierte Graphen zeigen als Muttersprachler/-innen, dass (2) die Strukturiertheit der Graphen bei Lerner/-innen im Erwerbsverlauf insgesamt ansteigt – die Graphen also zum Ende der Entwicklung stärker strukturiert sind als am Anfang – und dass (3) die Strukturiertheit im Erwerbsverlauf zunächst ab- und dann wieder zunimmt (U-Kurven-Verlauf). Shadrova (2020, Kap. 4) zeigt hierfür, dass eine statistische Analyse basierend auf Oberflächenfrequenzen und dem lexikalischen Assoziationsmaß DeltaP (Gries 2013) zwar einige interessante Einblicke, aber hinsichtlich der Struktur keine interpretierbaren Ergebnisse bringt. Ist die Graphanalyse hierin erfolgreicher?

3.3 Ergebnisse

Tatsächlich zeigt Abbildung 8, dass die Ergebnisse der Graphanalyse den Hypothesen in den Punkten (1) und (2) entsprechen, und dass für die belarussischen Lerner/-innen auch ein U-Kurven-Verlauf wie in Hypothese (3) angenommen sicht-

bar wird. Für die chinesischen Lerner/-innen ist das nicht der Fall.[13] Die beiden Gruppen repräsentieren unterschiedliche Vorfilterungen der enthaltenen Lexeme: In in den „Argumenten einschl. Subjekte“ sind alle Verben und ihre objektartigen Argumentlexeme enthalten (Prädikative, Akkusativ- Dativ-, Gentifivobjekte, Subjekte) sowie die Nominalergänzungen zu Präpositionalergänzungen, nicht jedoch die Präposition selbst. Nicht enthalten sind Attribute, Adverbien und nicht-objektartige Präpositionalphrasen. Der Graphtyp „Argumente ohne Subjekte“ enthält außerdem keine aktivitischen Subjekte (passivische Subjekte wurden als Akkusativobjekt für eine Gleichbehandlung von Strukturen wie *Man verwöhnt die Kinder* und *Die Kinder wurden verwöhnt* erhalten). Die Gründe dafür sowie der Vergleich mit anderen Graphspezifitäten sind ausführlich in Shadrova (2020, Kap. 3, 4 und 6) besprochen.

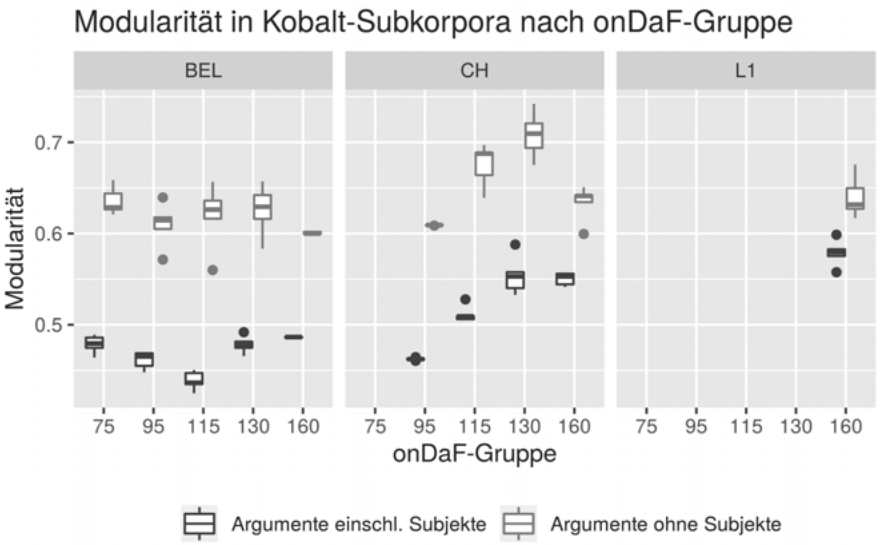

Abb. 8: Louvain-Modularität in Samples von je 10 Texten aus Kobalt-DaF-Subkorpora. Gruppierung nach onDaF-Testergebnis

13 Da Louvain-Modularität heuristisch berechnet wird, wurde für die Analyse auf Samples von je zehn Texten aus den Subkorpora — die Samples überlappen somit teilweise oder vollständig — in 350 Iterationen die Louvain-Modularität berechnet und das jeweilige Maximum als Ergebnis genommen. Die notwendige Zahl der Iterationen wurde vorher abgeschätzt, siehe Shadrova (2020, Appendix A1). Da die Datenpunkte (Samples) nicht unabhängig voneinander sind, ist eine Signifikanzberechnung nicht möglich.

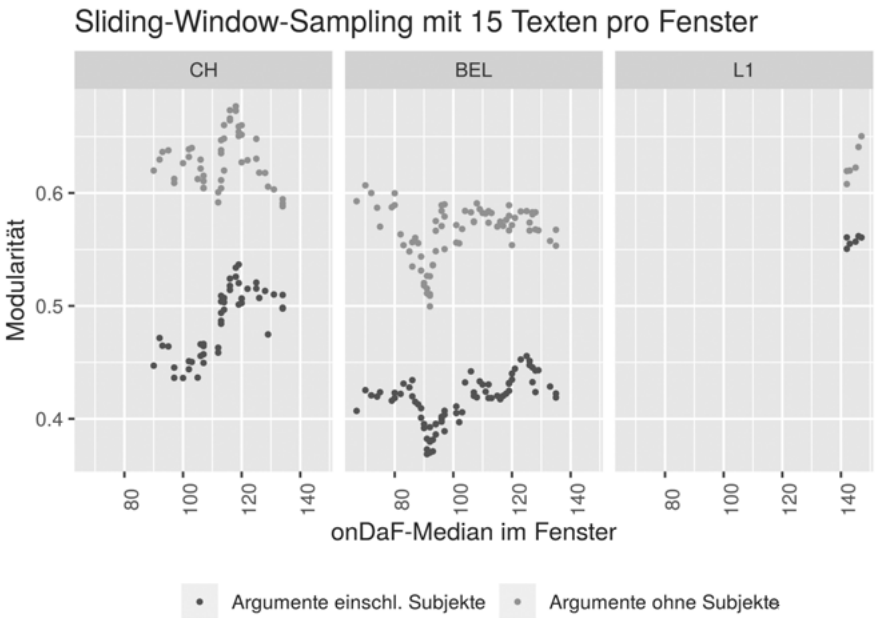

Abb. 9: Louvain-Modularität in Kobalt-DaF, Sliding-Window-Sampling: Ein Datenpunkt entspricht einem Ausschnitt von 15 Texten, sortiert nach onDaF-Rängen

In dieser Analyse sind Deutschlernende anhand ihrer Testergebnisse im validierten C-Test onDaF (heute onSet; Eckes 2010) in Gruppen eingeteilt worden. Die Zahlen geben dabei (exklusiv) die Punktegrenze an, sprich die Gruppe 95 enthält Texte, deren Autor/-innen zwischen 75 und 94 Punkten erreichten, die Gruppe 115 die von Autor/-innen mit 95–114 Punkten. Die onDaF-Gruppen wurden hier opportunistisch zum Erhalt möglichst vergleichbarer Gruppengrößen gewählt, entsprechen aber wohl ungefähr auch der Einteilung in die CEFR-Stufen A2 (nur BEL) – B1–B2.1 – B2.2 – C1(+). Für eine ausführliche Diskussion siehe Shadrova (2020, Kap. 3). Da es berechtigte Bedenken bzgl. der Einteilung in Stufen und des (diskreten) Gruppenvergleiches bei einer angenommen annähernd kontinuierlichen Punkteskala geben kann, zeigt Abbildung 7 dieselbe Analyse in einem sogenannten Sliding-Window-Sampling. Dafür wurden jeweils die Texte 1–15, 2–16, 3–17 usw. nach Rang in der onDaF-Punkteverteilung der jeweiligen Sprachgruppe als Subkorpus gewählt. Die Texte der 14 am weitesten fortgeschrittenen Lerner/-innen sind somit nur einmal in der Analyse berücksichtigt, die der Ränge 1-(n-15) jeweils 15-mal. Auch diese Analyse bestätigt die Ergebnisse und zeigt, dass es zwar eine Varianz zwischen den einzelnen Datenpunkten gibt, diese aber durchaus in er-

warteter Weise mit dem onDaF-Ergebnis korreliert. Wäre das nicht der Fall, wäre das Kurvenverhalten weitaus erratischer.[14]

Insgesamt zeigt die Graphanalyse somit in überraschender Klarheit eine Abhängigkeit der lexikalischen Struktur vom Erwerbsstand. Deutlich wird auch, dass sogar bei Muttersprachler/-innen eine vergleichsweise hohe Varianz besteht, sie aber auch in der niedrigeren Variante stärker strukturierte Graphen aufweisen als die Lerner/-innen.

Besonders interessant ist hier der Unterschied zwischen den chinesischen und belarussischen Lerner/-innen, der aus dem minutiösen Vergleich einzelner Koselektionen in der Weise wohl nicht quantitativ sichtbar geworden wäre. Über die Gründe lässt sich zunächst nur post-hoc spekulieren, es sei aber angemerkt, dass es hier interessante Parallelen zu typologischen Eigenschaften insbesondere des Chinesischen gibt, wo sogenannte Verb-Nomen-Komposita einen großen Teil der Verblexik ausmachen (Bodomo/So-sum/Che 2017; Badan 2013). Hierbei handelt es sich um Wörter, die einer Verbargumentstruktur gleichen, etwa pǎobù, 跑步 ‚Schritte laufen' oder kànshu, 看书 ‚Buch angucken', die aber eine reine Verbsemantik haben, nämlich resp. *laufen* und *lesen*. Mit dieser Prävalenz in der L1 erscheint es plausibel, dass auch in der L2 Mappings von Verb-Nomen-Kombinationen auf Tätigkeiten gelernt werden, wodurch der Reorganisationsprozess des Lexikons eingeschränkt oder ganz durchbrochen werden mag. Interessant ist hier eine Parallele zum chinesischen Erstspracherwerb, in dem Kinder, anders als im Erstspracherwerb der beschriebenen europäischen Sprachen, keinen sogenannten Nomenbias zeigen (Tardif 1996; Tardif/Shatz/Naigles 1997). Soll heißen: Chinesischsprachige Kinder lernen Verben und Nomen gleichermaßen leicht, während Kinder in bestimmten Phasen des Spracherwerbs vieler europäischer Sprachen weniger, später, und weniger leicht Verben lernen. Andere mögliche Gründe liegen in kulturellen und pädagogischen Kontexten sowie in der unterschiedlichen Bearbeitung derselben Aufgabe: Belarussische Lerner/-innen produzieren allgemein weniger analytische, sondern stärker erzählende oder beschreibende Texte, was mit einer anderen Koselektionsstruktur korrelieren könnte.

14 Shadrova (2020, Kap. 6) diskutiert hierzu umfassend verschiedene Validierungs- und Samplingfragen sowie die Interpretation der abfallenden Kurve gegen Ende des Punkteverlaufs. Tatsächlich scheint es sich hier um einen Textstruktureffekt zu handeln, nach dem die Modularität bei besonders langen Texten bei den belarusischen Lerner/-innen zum Ende hin sinkt bzw. am Ende niedriger liegt als am Anfang der Texte, und diese zugleich zunehmend längere Texte schreiben.

Eine Verifikation dieser Ergebnisse erfordert zunächst die Replikation auf anderen Korpora unter Berücksichtigung der neuen Hypothese (bspw. typologische Unterschiede und Unterschiede nach Register). Während in der Analyse eine umfassende interne Validierung anhand verschiedener Samplingmethoden durchgeführt wurde, die auf eine Robustheit der Ergebnisse für dieses spezielle Korpus deutet, lässt sich aus dieser Tatsache allein noch kein Rückschluss auf eine Allgemeingültigkeit oder allgemeine Nützlichkeit der Methode ziehen. Wichtig ist jedoch, dass eine Replikation prinzipiell möglich ist, weil die Datenmenge – 171 Texte einschließlich der muttersprachlichen Vergleichstexte – auch im Rahmen kleinerer Forschungsprojekte überschau- und vor allem technisch und linguistisch beherrschbar ist. Konkret erfordert das Projekt zwar einiges an Rechenkapazität, ist aber mit gewissen Einschränkungen noch auf einem leistungsstarken Laptop oder einem institutsüblichen Server realisierbar. Wichtiger ist jedoch, dass der Aufwand für die manuelle Korrektur der Dependenzparses – hier MaltParser (Nivre/Hall/Nilsson 2006) mit der Dependenzgrammatik von Foth (2006) – noch im Bereich des Leistbaren liegt. Da die Parser-Accuracy insbesondere bei den Lernertexten der unteren und mittleren Erwerbsstufen sehr niedrig ist, hängt die Durchführbarkeit einer solchen Analyse letztlich hauptsächlich an dieser Variable. Sowohl die Rechen-, als auch die Annotationskapazität geraten hingegen schnell in unbeherrschbare Bereiche, wenn man auf größere Korpora setzt. Somit ist die Korpusgröße hier ein limitierender Faktor, bildet anders als bei frequentistischen statistischen Analysen jedoch eine *obere* Grenze – eine Erweiterung der quantitativen Perspektive auch für diejenigen, die mit elizitierten oder intrinsisch beschränkten, etwa historischen Daten arbeiten.

Bezogen auf unsere Hypothesen 1.2.1–1.2.3 ergibt sich, dass (3) „kleine bis mittelgroße Korpora erlauben eine tiefe, linguistisch informierte Analyse“ zwingend erfüllt ist, denn die hier vorgestellte Analyse wäre in größeren Korpora nicht umsetzbar. (1) „Linguistische Analysen und Hypothesen (hier zum Spracherwerb) lassen sich nicht auf Oberflächenphänomene reduzieren“ und (2) „abstrakte sprachliche Phänomene erfordern eine saubere, angemessene, theoretische Modellierung“ sind hier dialektisch mit der Methode verflochten: Koselektion ist in der empirischen Linguistik bislang überwiegend statistisch anhand von Exemplaren betrachtet worden, aber das liegt nicht daran, dass das linguistische Modell *eigentlich* eine stochastische Perspektive vorsieht, sondern am Anspruch, das Phänomen quantitativ, graduell, und über-lexikografisch, also nicht nur auf die Qualität der einzelnen Wortpaare hin, zu fassen. Bei näherer Betrachtung stellt sich das stochastische Modell jedoch als problematisch heraus, insbesondere dort, wo die eigentlichen Exemplare dann doch wieder gruppiert, nämlich als Ausdruck einer Gesamttendenz wie des idiomatischen Prinzips verstanden werden sollen. Ist das der Fall, verschwindet der Zugang über die reinen Oberflächenfor-

men (Hypothese 1.2.1), was dann eine komplexe Modellierung erfordert (1.2.2), die aber nur im Rahmen geeigneter Methodenframeworks und Modelle überhaupt empirisch umsetzbar ist. Die Entwicklung und umfassende Validierung dieser Methoden wiederum erfordert ein tiefes Verständnis der Daten, da sonst die Gefahr besteht, Zugeständnisse an das quantitative Modell zu machen, die nicht linguistisch motiviert oder motivierbar sind. In der Konsequenz ergibt sich, dass man die Konkretheit der Daten sehr genau kennen muss, um die Abstraktion des Koselektionsphänomens überhaupt linguistisch sinnvoll beschreiben zu können, und dass die (quantitative) Gradierbarkeit erst auf Grundlage einer adäquaten und differenzierten qualitativen Beschreibung erfassbar wird.

4 Rhetorische Strukturen

4.1 Zielsetzung

Um einen kohärenten, logischen und klar strukturierten registerangemessenen Text in einer Fremdsprache zu verfassen, benötigt man nicht nur lexikalische und grammatische Kenntnisse, sondern auch Kenntnisse über die registerspezifische Rhetorik/Textstruktur der Zielsprache. In unserem letzten kurzen Beispiel möchten wir nun auf rhetorische Strukturen[15] zu sprechen kommen und wieder alle drei Hypothesen betrachten. Wir wissen aus vielen Arbeiten, dass ihre L1 die Sprache von Lernenden auf allen Ebenen beeinflusst.[16] Dies ist relativ leicht zu sehen an lexikalischen und oberflächennahen syntaktischen Phänomenen, aber schwieriger auf abstrakteren ‚höheren' Ebenen. In Abschnitt 3 wurde bereits gezeigt, dass die L1 einen Einfluss auf den Erwerb von Koselektionsbeschränkungen hat. Wie schon in der Einleitung erwähnt, ist es auch oft so, dass man bestimmten Texten ihre Nichtmuttersprachlichkeit anmerkt, auch wenn sie keine Fehler enthalten. Das kann manchmal daran liegen, dass die Texte rhetorisch unterschiedlich strukturiert sind. Hier betrachten wir die rhetorischen Struk-

15 Andere Termini sind z. B. Kohärenzrelationen und Diskursrelationen (Stede 2018).

16 Transfer (oder manchmal auch Interferenz) ist komplex und nicht immer einfach zu modellieren (Ellis 2006; Jarvis 2010; Osborne 2015). Das Phänomen, welches wir hier untersuchen, ist stark kulturell geprägt und wird durch Schulunterricht mitbestimmt — manche Autor/-innen würden es daher nicht als Transfer bezeichnen. Daher sprechen auch wir hier eher von L1-Einfluss.

turen von chinesischen L1-Sprecher/-innen im Vergleich mit denen deutscher L1-Sprecher/-innen.[17] Wir möchten das an den Beispielen (3a) und (3b) aus dem Kobalt-Korpus illustrieren. Es handelt sich um die ersten Sätze aus einem Muttersprachler- (3a) und einem Lernertext mit L1 Chinesisch (3b), die jeweils die Aussage „Geht es Jugend heute besser als früher?“ diskutieren.

(3a) Der heutigen Jugend geht es besser als früheren Generationen. Dieser Aussage würde ich nur bedingt zustimmen, denn meiner Meinung nach gibt es für beide Generationen Vor- und Nachteile. (Kobalt-DaF-L1-Korpus, Text DEU_003_2012_04; https://hu.berlin/bsp3a)

(3b) Seit der Reform- und Öffnungspolitik im Jahr 1978 gibt es in China große Veränderungen, besonders die Wirtschaft entwickelt sich sehr schnell und der Lebensstandard wird höher Jahr für Jahr. Viele Jugendlichen glauben, dass sie in einer guten Zeit leben. (Kobalt-DaF-L2-Korpus, Text CMN_012_2012_03; https://hu.berlin/bsp3b)

Die Unterschiede in der rhetorischen Struktur sind bereits in den ersten Zeilen offensichtlich. Während in (3a) gleich eine eigene Meinung präsentiert und diese dann begründet wird, beginnt der oder die Deutschlernende in (3b) einleitend mit einer generellen politischen Einordnung. Das wäre nicht weiter interessant, wenn es sich hier um idiosynkratische Texte mit verschiedenen Schreibanlässen handeln würde. Es handelt sich jedoch um Texte derselben Textsorte, desselben Themas und desselben Schreibanlasses. Mit Blick auf die genannten zwei Kohorten des Kobalt-DaF-Korpus (L1-Chinesisch- und Deutsch-Schreiber/-innen) soll ermittelt werden, ob die offensichtlichen Unterschiede in (3a) und (3b) systematisch gelten. Allgemeiner formuliert: Wir wollen ermitteln, ob sich die rhetorischen Strukturen bei Deutschlerner/-innen und denen von deutschen Muttersprachler/-innen unterscheiden, und inwieweit die Unterschiede (wenn es sie gibt) auf eine kulturelle Prägung und den L1-Hintergrund zurückzuführen sind.

Wir erachten den Vergleich zwischen chinesischen L1-Sprecher/-innen und deutschen L1-Sprecher/-innen als besonders interessant, weil sich hieran Rhetorik als sprach- und kulturspezifisches Phänomen (Kaplan 1966) untersuchen lässt. Chinesische L1-Sprecher/-innen und deutsche L1-Sprecher/-innen unterscheiden sich nicht nur einfach in ihrer Muttersprache, sondern auch in ihrer Zugehörigkeit zu verschiedenen Kulturkreisen. Es wurde bereits gezeigt, dass sich gewisse

17 Das Wissen über rhetorische Strukturen stellt einen Teil des Registerwissens von Lerner/-innen dar.

Schreibkonventionen des Chinesischen und Deutschen unterscheiden (Liao/Chen 2009; Tsao 1983; Liu 2005; Lehker/Wang 2006). Hieran wollen wir mit der folgenden Untersuchung anschließen.

Unsere spezifischen Forschungsfragen beziehen sich sowohl auf formale als auch auf funktionale Aspekte:

a) Wie unterscheidet sich die Verwendung bestimmter rhetorischer Relationen zwischen den beiden Gruppen?
b) Wie werden diese Relationen sprachlich realisiert?
c) Welche Einflussfaktoren bedingen die Unterschiede, die dabei ermittelt wurden?

Die Antworten auf diese Fragen können dabei helfen, a) die Gemeinsamkeiten und Unterschiede bezüglich der rhetorischen Strategien von chinesischen L1-Sprecher/-innen sowie von deutschen L1-Sprecher/-innen zu erkennen; b) L1-Einfluss in Bezug auf rhetorische Struktur zu entdecken; c) als Referenz dienen, um die Lehr- und Lernmethodik des DaF-Unterrichts in China zu verbessern.

4.2 Datenauswertung

Um diese Fragen zu beantworten, nutzen wir wieder das Kobalt-Korpus. Das Kobalt-DaF Korpus eignet sich sowohl aus der Perspektive der Menge der Aufsätze (siehe Tab. 1), welche systematische Analysen und eine manuelle Datenbearbeitung gleichzeitig ermöglicht, als auch aus der Sichtweise der Aufgabenstellung für unser Forschungsziel. Wie bereits erwähnt, bearbeiten alle Teilnehmer/-innen in Kobalt dieselbe Fragestellung („Geht es der Jugend heute besser als früheren Generationen?“). Alle Aufsätze wurden unter denselben Rahmenbedingungen verfasst. Dies ist für die Studie von großer Bedeutung, weil die Aufgabenstellung Einfluss auf die Schreibstrategien und insofern auch auf die rhetorischen Strukturen der Texte hat (Ferretti/Lewis/Andrews-Weckerly 2009).

Für die vorliegende Studie wurden 20 argumentative Texte von chinesischen Deutschlerner/-innen und 20 argumentative Texte von deutschen L1-Sprecher/-innen verwendet, die im Kobalt-Korpus zur Verfügung stehen (vgl. Tab. 1).

Tab. 1: Verteilung der für diese Studie verwendeten Texte

L1	Genre	Anzahl der Texte	Textlänge i.D.	Token
Chinesisch	argum. Texte	20	521	13.949
Deutsch	argum. Texte	20	488	12.984

Statt die Originaltexte anzuschauen, berücksichtigen wir bei der Annotation von rhetorischen Strukturen die bereits vorliegenden Parallelfassungen der Zielhypothese, da mögliche Fehler nicht Forschungsgegenstand dieser Arbeit sind, sondern im Gegenteil das Verständnis für die Aufsätze behindern können. Besonders erwähnenswert ist, dass nicht nur die Texte der chinesischen Deutschlerner/-innen, sondern auch die Vergleichstexte der deutschen Muttersprachler/-innen mit Zielhypothesen annotiert wurden. Der Grund dafür liegt darin, dass die L1-Vergleichsdaten in vielerlei Hinsicht nicht der Norm entsprechen müssen (vgl. Granger 2015).

Das Kobalt-Korpus wurde bereits ausgiebig annotiert. Um die rhetorischen Strukturen der Texte eingehend zu analysieren, benötigen wir allerdings noch eine zusätzliche Annotationsebene für die Beschreibung rhetorischer Strukturen. Dazu wird das funktionale Framework der *Rhetorical Structure Theory* (RST; Mann/Thompson 1988) für die weitergehende Annotation und Auswertung verwendet. Als eine der beliebtesten Theorien im Bereich der Diskursanalyse bietet RST das Beschreibungsinventar dafür, die rhetorische Struktur und somit die Organisation eines Texts durch eine Baumstruktur darzustellen. Die Blätter des Baums werden als Elementary Discourse Units (EDUs) bezeichnet. Eine EDU entspricht einer selbstständigen Sprechhandlung (Stede (Hg.) 2016). Die EDUs oder die rekursiv größeren Textabschnitte werden dann durch rhetorische Relationen verbunden. In unserer Studie werden insgesamt 33 rhetorische Relationen auf der Grundlage der Annotationsrichtlinie des Potsdamer Kommentarkorpus (PCC),[18] die speziell für deutschsprachige argumentative Texte entwickelt wurde, verwendet. Nach ihrer inhaltlichen und strukturellen Unterscheidung gliedern sich diese Relationen in vier Kategorien (Tab. 2). Primär pragmatische Relationen beschreiben die Argumentation bzw. das Verhalten des Autors oder der Autorin. Primär semantische Relationen kommen zum Einsatz, wenn Sachverhalte neutral beschrieben werden. Textuelle Relationen erfüllen einen organisierenden Zweck. Multinukleare Relationen sind im Gegensatz zu den anderen drei Relationstypen nicht solche zwischen einer übergeordneten (Nukleus) und einer untergeordneten (Satellit) EDU, sondern Relationen zwischen gleichrangigen EDUS (bzw. zwei Nuklei).

18 http://angcl.ling.uni-potsdam.de/resources/pcc.html (Stand: 23.9.2020).

Tab. 2: Überblick der rhetorischen Relationen nach leicht modifizierten Annotationsrichtlinien für das PCC-Korpus[19]

Primär pragmatische Relationen	Primär semantische Relationen	Textuelle Relationen	Multinukleare Relationen
Background	Circumstance	Preparation	Contrast
Antithesis	Condition	Restatement	Sequence
Concession	Otherwise	Summary	List
Evidence	Unless		Conjunction
Reason	Elaboration		Joint
Reason-N	E-Elaboration		
Justify	Interpretation		
Evaluation-S	Means		
Evaluation-N	Cause		
Motivation	Result		
Enablement	Purpose		
Question	Solutionhood-N		
	Solutionhood-S		

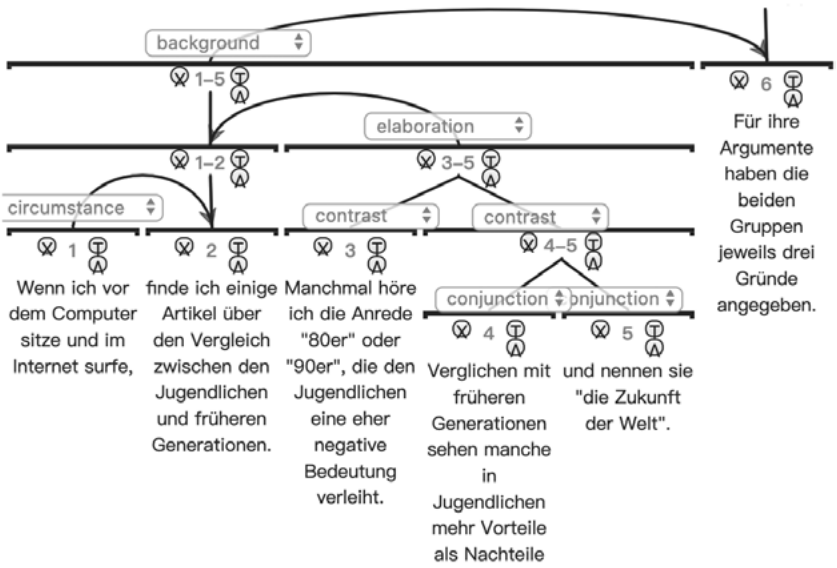

Abb. 10: Ein RST-Analysebeispiel (Textauszug aus Kobalt_CMN_017_2012_04)

19 Die dort formulierten Annotationsrichtlinien enthalten 31 rhetorische Relationen. Im Laufe des iterativen Annotationsvorgangs wurden lediglich zwei differenzierende Relationen („Question“ und „Solutionhood- S“) ergänzt.

Die Annotation auf Diskursebene ist eine extrem zeit- und ressourcenintensive Aufgabe (vgl. Gries/Berez 2017 zur Einordnung dieses Annotationstyps). Obwohl es schon einige Versuche gibt, die Annotation der rhetorischen Struktur semiautomatisch zu verwirklichen, werden die meisten Annotationsarbeiten bezüglich rhetorischer Strukturen aufgrund der unvermeidlichen Subjektivität, Komplexität und Mehrdeutigkeit des Diskurses vorwiegend manuell durchgeführt. Der RST-Analyse ist eine gewisse Subjektivität inhärent und wird auch als Rekonstruktion der Absicht der Autorin oder des Autors aus der Rezipientensicht betrachtet (Stede (Hg.) 2016). Damit die Annotation trotz dieser Einschränkungen nachvollziehbar und zuverlässig ist, wird unsere Annotationsarbeit sorgfältig nach einem ausführlichen Annotationsplan durchgeführt. Zwei fachlich ausgebildete Linguistinnen werden als Annotatorinnen eingesetzt. Zum Schluss wird die Qualität der Annotationen aller Texte durch eine systematische Berechnung des Inter-Annotator Agreements evaluiert. Mithilfe des Tools RST-Tace (Wan et al. 2019) ist die Annotation aller 40 argumentativen Texte von den beiden Annotatoren mit dem Kappa-Wert 0,74 beachtlich übereinstimmend. Abbildung 10 zeigt ein Beispiel der Annotation eines Textauszugs aus dem Kobalt-Korpus, visualisiert mithilfe des Werkzeugs RSTWeb (Zeldes 2016).

4.3 Ergebnisse

Wesentliche Ergebnisse der Studie sind in Abbildung 11 dargestellt.

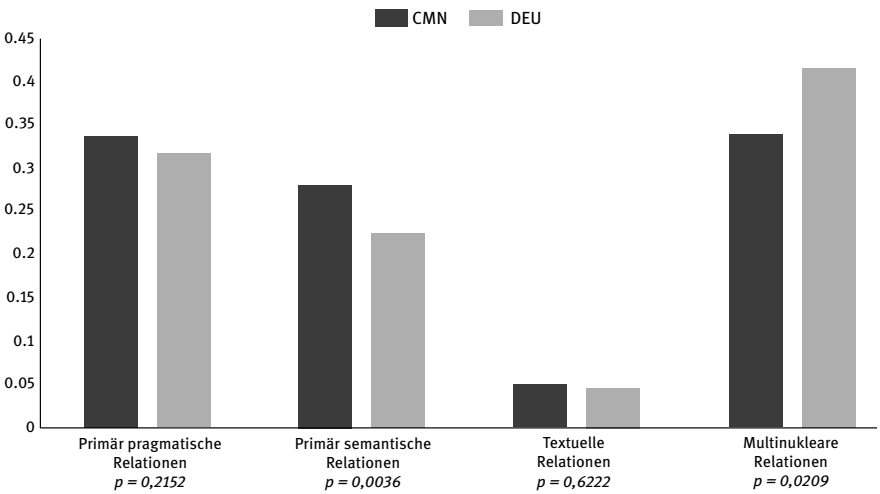

Abb. 11: Verteilung der rhetorischen Strukturen in den vier in Tabelle 2 genannten Kategorien

Gemäß Abbildung 11 ist ersichtlich, dass sich die deutschen Muttersprachler/-innen und die chinesischen Deutschlerner/-innen in den Kategorien der primär pragmatischen Relationen sowie der textuellen Relationen kaum unterscheiden. Allerdings verwenden die deutschen Muttersprachler/-innen signifikant mehr multinukleare Relationen als die chinesischen Deutschlerner/-innen (χ^2-Test, $p \approx 0{,}02$) und die chinesischen Deutschlerner/-innen signifikant mehr primär semantische Relationen als die Vergleichsgruppe (χ^2-Test, $p \approx 0{,}004$).

Gemäß unseren Annotationsrichtlinien gehören die rhetorischen Relationen Contrast, Sequence, List, Conjunction und Joint zu den multinuklearen Relationen. Diese Relationen werden häufig verwendet, wenn nebeneinanderstehende oder gegensätzliche Behauptungen bzw. Argumente, die eine gemeinsame Rolle für die Textfunktion spielen, ausführen sollen. Ein typisches Lernerbeispiel für die Verwendung verschiedener Äußerungen in dieser nebengeordneten Weise ist in Abbildung 12 illustriert: Die Äußerungen *Erstens bekommt man heute mehr Informationen als früher...*, *Zweitens ist die Jugend heute wegen neuer Lebensformen eifriger...* und *Drittens nimmt die Jugend mehr Kontakt mit der Außenwelt beim Reisen und Studieren auf...* bilden eine Reihe, die mit der Kategorie „List“ annotiert wird.

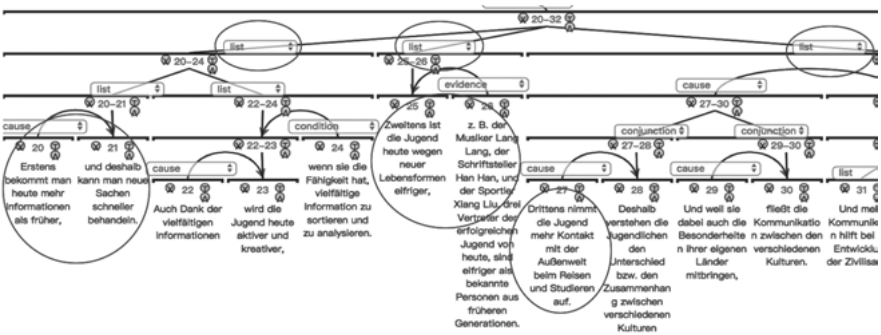

Abb. 12: Beispiel der rhetorischen Relation „list“ im Lernertext Kobalt_CMN_006_2012_04

Die Kategorie der primär semantischen Relationen bilden einen Sachverhalt mit einer hierarchischen Bedeutungsstruktur ab, bei der einem übergeordneten Äußerungsteil (dem Nukleus) ein weiterer Äußerungsteil mit Bedeutungsrelationen wie „Cause“ (sachlicher Grund), „Condition“ (sachliche Bedingung) oder „Elaboration“ (nähere Erläuterung) untergeordnet wird. Die deutlich vermehrte Verwendung der Relationen in dieser Kategorie lässt schlussfolgern, dass von den chinesischen Deutschlerner/-innen Sachverhalte und Ereignisse neutraler dargestellt

werden als von den deutschen Muttersprachler/-innen. Dies lässt sich durch die Analyse spezifischer Relationen eingehend abklären.

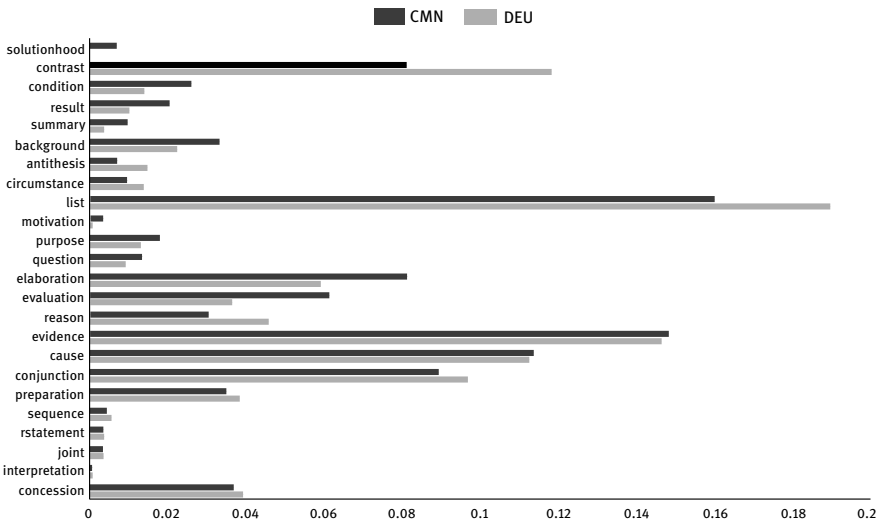

Abb. 13: Die Verwendungshäufigkeit der rhetorischen Relationen in den verglichenen Gruppen

Im Hinblick auf die einzelnen RST-Relationen, die in Abbildung 13 dargestellt sind, ist es interessant zu sehen, dass beide Gruppen die Relationen „List“, „Evidence“, „Cause“, „Conjunction“, „Contrast“ und „Elaboration“ am häufigsten verwenden. Jedoch werden bei den chinesischen Deutschlerner/-innen gleichzeitig die Kategorien „List“, „Conjunction“ und „Contrast“ deutlich weniger verwendet. Bei „Elaboration“ und „Evidence“ ist die Relation umgekehrt. „Elaboration“ und „Evidence“ haben oft eine unterstützende Funktion, durch die Behauptungen bzw. Argumente der Autor/-innen ergänzt werden und detailliertere Informationen liefern, womit die Ausführungen klarer und überzeugender wirken. Dies lässt Folgendes schlussfolgern: Während die deutschen Muttersprachler/-innen mehr Argumente einführen und sie tendenziell knapper ausführen, neigen die chinesischen Sprecher/-innen eher dazu, weniger Argumente hervorzubringen, dafür aber jedes Argument ausführlicher zu belegen.

Darüber hinaus ist in den Daten zu sehen, dass die chinesischen Deutschlerner/-innen im Vergleich zu den deutschen Muttersprachler/-innen die Kategorie „Reason“ deutlich seltener verwenden; dies ist eine Relation, bei der subjektive Einschätzungen oder Behauptungen als unterstützende Argumente dienen. Eine Erklärung für dieses Phänomen könnte sein, dass die chinesischen Deutschlerner/

-innen eher Heckenstrategien[20] verwenden, statt die subjektive Haltung hervorzukehren. Sie versuchen, Distanz zu den Thesen aufzubauen, indem sie ihre eigene Meinung anderen Personen oder feststehenden Sachverhalten zuschreiben. Statt *meiner Meinung nach* schreiben die chinesischen Lernenden beispielweise Ausdrücke wie *Manche sind der Meinung* (Kobalt_CMN_010_2012_03). Eine systematische Untersuchung dieses Phänomens wird in Wan (i. Vorb.) durchgeführt.

In Hinblick auf die textuellen Strukturen haben wir auf der Grundlage der annotierten rhetorischen Relationen die Texte tiefergehend analysiert. Dabei wurde ersichtlich, dass die chinesischen Deutschlerner/-innen dazu neigen, ihre Texte mit allgemeinem Wissen wie der wirtschaftlichen oder politischen Entwicklung Chinas einzuleiten. Die Beispiele (4a) und (4b) stellen dies sehr anschaulich dar. Dieser Befund kann auch durch die umfassende Nutzung der rhetorischen Relation „Background" von den chinesischen Deutschlerner/-innen bestätigt werden. Interessant ist, dass in einer Studie über die Englischlerner/-innen (Callies 2015) gezeigt wurde, dass chinesische Englischlerner/-innen bei der Einführung in die Thesen ebenfalls überwiegend auf das allgemeine Weltwissen Bezug nehmen.

(4a) Mit der Entwicklung der Welt haben sich die Menschen sehr verändert. (Kobalt-DaF-L2-Korpus, Text CMN_001_2012_03; https://hu.berlin/bsp4a)

(4b) Seit der Reform- und Öffnungspolitik im Jahr 1978 gibt es in China große Veränderungen. (Kobalt-DaF-L2-Korpus, Text CMN_012_2012_03; https://hu.berlin/bsp4b)

Darüber hinaus wird in unserer Studie gezeigt, dass die chinesischen Deutschlerner/-innen ihre Texte untereinander sehr ähnlich strukturieren, während die deutschen Muttersprachler/-innen unterschiedlichere Textstrukturen verwenden.

Die Mehrheit der chinesischen Deutschlerner/-innen hält die folgende Reihenfolge beim Verfassen ihrer Texte ein:

Einführung in die These → positive und negative Haltungen → Meinung des Autors

Dieses Muster wird einerseits vorwiegend im DaF-Unterricht in China gelehrt und deshalb auch viel von den Lernenden realisiert. Andererseits kann diese Struktur

20 Heckenausdrücke (auf Englisch: hedging expressions) beziehen sich als ein Teil der epistemischen Modalität auf alle sprachlichen Mittel, die verwendet werden, um eine ausdrückliche Verpflichtung hinsichtlich der Wahrheit zu vermeiden (Hyland 1998).

auch auf die traditionelle Textstruktur des chinesischen qǐ-chéng-zhuǎn-hé[21] und bā-gǔ-wén[22] zurückzuführen sein. Inwieweit die traditionellen Schreibkonventionen auf das Schreiben im modernen Chinesischen und in Fremdsprachen beeinflussen, wird in Wan (i. Vorb.) tiefergehend untersucht.

Im Vergleich zu den relativ einheitlichen Textstrukturen der chinesischen Deutschlernenden variieren, wie gesagt, die Texte der deutschen Muttersprachler/-innen relativ stark. Es finden sich zwar ähnliche Muster wie bei den chinesischen L1-Sprecher/-innen, bei denen die eigene Meinung erst am Ende genannt wird, es werden aber auch viele weitere Muster wie die folgenden verwendet:

Meinung des Autors → Begründung → Zusammenfassung
Begrifflichkeitsklärung → Meinung des Autors → Begründung

Außerdem zeigt die Studie folgende Tendenzen: Während die deutschen Muttersprachler/-innen ihre Argumente nicht explizit strukturieren, teilen die chinesischen Deutschlerner/-innen ihre Argumente häufig in drei oder vier Hauptpunkte und ordnen sie explizit nach einer bestimmten Reihenfolge. Dabei können wieder Spuren der traditionellen chinesischen Schreibkonventionen gefunden werden:

Argumente werden entweder parallel (jeder Punkt stellt die betreffende These aus verschiedenen Perspektiven dar) oder progressiv (jeder Punkt baut aufeinander auf, um die These zu unterstützen) arrangiert.

Das folgende Beispiel (5) zeigt eine progressive Reihenfolge der Argumente in einem Text von einem bzw. einer chinesischen Deutschlernenden:

(5) Erstens sind die Jugendlichen jung und energisch [...] zweitens machten frühere Generationen auch Fehler [...] Drittens ist wohl der wichtigste Grund, dass die Jugendlichen formbar sind und Potential haben. (Kobalt-DaF-L2-Korpus, Text CMN_006_2012_03; https://hu.berlin/bsp5)

Zusammenfassend lässt sich Folgendes festhalten: Basierend auf einem kleinen, aber tief annotierten Korpus, hat diese Studie zu den rhetorischen Strukturen in Texten chinesischer Deutschlernender und deutscher Muttersprachler/-innen viele interessante Phänomene aufgezeigt, so z. B.:

21 qǐ-chéng-zhuǎn-hé: Thematisierung → Entwicklung mit positiver Interpretation → Übergang zu einem anderen Standpunkt → Schlussfolgerung (Yang/Yang 2010).
22 bā-gǔ-wén: Öffnung → Erweiterung → vorläufige Exposition → erster Punkt → zweiter Punkt → dritter Punkt → Schlusspunkt → Schlussfolgerung (Cai 1993).

a) Die chinesischen Deutschlernenden neigen dazu, ihre Texte rhetorisch einheitlicher zu organisieren als die Muttersprachler/-innen.
b) Die chinesischen Deutschlernenden verwenden weniger Argumente im Vergleich zu den deutschen Muttersprachler/-innen, aber führen jedes Argument detaillierter aus.
c) Die chinesischen Deutschlernenden tendieren dazu, im Gegensatz zur Gruppe der Muttersprachler/-innen ihre eigene, „versachlichte" Meinung am Ende des Textes zu positionieren.

Diese Tendenzen zeigen klar, dass trotz gleichem Register, gleichen Schreibanlässen, Textproduktionskontexten und vielen inhaltlichen Übereinstimmungen erhebliche Abweichungen in der Textorganisation vorliegen können. Dies motiviert zu tiefergehender textlinguistischer Forschung, der Überprüfung alter und der Entwicklung neuer Textstrukturtheorien sowie der Verwobenheit sprachlichen Handelns mit soziokulturellen Aspekten.

5 Zusammenfassung und Ausblick

Wir haben an drei Beispielen – der Auswertung von Modifikatoren, koselegierten Einheiten und rhetorischen Strukturen in Texten fortgeschrittener Lernender des Deutschen als Fremdsprache – gezeigt, dass viele Forschungsfragen erfordern, von der Textoberfläche zu abstrahieren, präzise linguistische Modelle zu formulieren und diese auf die zu untersuchenden Sprachdaten anzuwenden. Die drei Forschungsanliegen haben zunächst drei vollkommen verschiedene Klassifikationen und Datenarchitekturen hervorgebracht. Die Forschungsergebnisse konnten nur dadurch erzielt werden, dass

- ein wesentlicher Teil der gesamten Forschung in die Erstellung geeigneter, wissenschaftlich valider linguistischer Modellierungen investiert wurde,
- die primären Textdaten der DaF-Lernenden flexibel weiterverarbeitet werden und die entworfenen linguistischen Modelle auf sie angewendet werden konnten,
- die Gesamtheit der Daten noch so überschaubar war, dass gewisse tiefe Analysen manuell durchgeführt werden konnten.

Unsere Studien haben auf allen drei Analyseebenen spracherwerbsspezifische Effekte nachgewiesen, die zeigen, dass Sprachenlernen alles andere als nur der Erwerb einer Wortschatzgrundlage und grammatischer Regeln ist, sondern dass Spracherwerb durch viel mehr Herausforderungen bestimmt ist. Dies bestätigt die eingangs formulierten Annahmen 1.2.1–1.2.3.

Literatur

Aijmer, Karin (2002): Modality in advanced Swedish learners' written interlanguage. In: Granger, Sylviane/Hung, Joseph/Petch-Tyson, Stephanie (Hg.): Computer learner corpora, second language acquisition and foreign language teaching. (= Language Learning & Language Teaching 6). Amsterdam/Philadelphia: Benjamins, S. 55–76.

Alexopoulou, Theodora/Michel, Marije/Murakami, Akira/Meurers, Detmar (2017): Task effects on linguistic complexity and accuracy: a large-scale learner corpus analysis employing natural language processing techniques. In: Language Learning 67, S1 (Special Issue: Currents in Language Learning Series: Experimental, Computational, and Corpus-Based Approaches to Language Learning: Evidence and Interpretation), S. 180–208.

Badan, Linda (2013): Verb-object constructions in Mandarin: a comparison with Ewe. In: The Linguistic Review 30, 3, S. 373–422.

Bahns, Jens (1993): Lexical collocations: a contrastive view. In: ELT journal 47, 1, S. 56–63.

Ballier, Nicholas/Martin, Philippe (2015): Speech annotation of learner corpora. In: Granger/Gilquin/Meunier (Hg.), S. 107–134.

Bartsch, Sabine (2004): Structural and functional properties of collocations in English: a corpus study of lexical and pragmatic constraints on lexical co-occurrence. Tübingen: Narr.

Blondel, Vincent D./Guillaume, Jean-Loup/Lambiotte, Renaud/Lefebvre, Etienne (2008): Fast unfolding of communities in large networks. In: Journal of Statistical Mechanics: Theory and Experiment 10, S. 1–12.

Bodomo, Adams/Yu, So-sum/Che, Dewei (2017): Verb-object compounds and idioms in Chinese. In: Mitkov, Ruslan (Hg.): Computational and corpus-based phraseology. Second international conference, Europhras 2017, London, UK, November 13–14, 2017. Proceedings. (= Lecture Notes in Computer Science 10596). Cham: Springer, S. 383–396.

Brainerd, Barron/Chang, Sun Man (1982): Number of occurrences in two-state Markov chains, with an application in linguistics. In: The Canadian Journal of Statistics/La revue canadienne de statistique 10, 3, S. 225–231.

Cai, Guanjun (1993): Beyond "bad writing": teaching english composition to Chinese ESL students. Konferenzpapier. In: Annual meeting of the conference on college composition and communication (44th, San Diego, CA, March 31–April 3, 1993), S. 1–24.

Callies, Marcus (2015): Towards a function-driven approach to annotating learner corpora: the case of topic marking. In: Learner corpus research conference (LCR 2015), 11–13 September, Nijmegen, Netherlands. Nimwegen: Radboud Universität, S. 8–11.

Collins, Harry (2016): Reproducibility of experiments: experimenters' regress, statistical uncertainty principle, and the replication imperative. In: Atmanspacher, Harald/Maasen, Sabine (Hg.): Reproducibility: principles, problems, practices, and prospects. Hoboken: Wiley, S. 65–82. DOI: 10.1002/9781118865064.ch4.

Council of Europe (2017): Common European framework of reference for languages: learning, teaching, assessment. Companion volume with new descriptors. Provisional edition. Straßburg: Council of Europe. Internet: www.coe.int/lang-cefr (Stand: 29.9.2020).

Croft, William (2001): Radical construction grammar. Syntactic theory in typological perspective. Oxford: Oxford University Press.

Dörnyei, Zoltán (2017): Conceptualizing learner characteristics in a complex, dynamic world. In: Ortega, Lourdes/Han, ZhaoHong (Hg.): Complexity theory and language development. In celebration of Diane Larsen-Freeman. (= Language Learning & Laguage Teaching 48). Amsterdam/Philadelphia: Benjamins, S. 79–96.

Eckes, T. (2010): Der Online-Einstufungstest Deutsch als Fremdsprache (onDaF): Theoretische Grundlagen, Konstruktion und Validierung. In: Grotjahn, Rüdiger (Hg.): Der C-Test: Beiträge aus der aktuellen Forschung/The C-test: Contributions from current research. (= Language Testing and Evaluation 18). Frankfurt a. M.: Lang, S. 125–192.

Ellis, Nick C. (2006): Selective attention and transfer phenomena in L2 acquisition: contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. In: Applied Linguistics 27, 2, S. 164–194.

Evert, Stefan/Uhrig, Peter/Bartsch, Sabine/Proisl, Thomas (2017): E-VIEW-alation – a large-scale evaluation study of association measures for collocation. In: Kosem, Iztok/Tiberius, Carole/Jakubíček, Miloš/Kallas, Jelena/Krek, Simon/Baisa, Vít (Hg.): Electronic lexicography in the 21st century: lexicography from scratch. Proceedings of eLex 2017 conference. Leiden, the Netherlands, 19–21 September 2017. Brünn: Lexical Computing CZ s.r.o., S. 531–549.

Feest, Uljana (2016): The experimenters' regress reconsidered: replication, tacit knowledge, and the dynamics of knowledge generation. In: Studies in History and Philosophy of Science Part A 58, S. 34–45.

Ferretti, Ralph P./Lewis, William E./Andrews-Weckerly, Scott (2009): Do goals affect the structure of students' argumentative writing strategies? In: Journal of Educational Psychology 101, 3, S. 577–589.

Foth, Kilian A. (2006): Eine umfassende Constraint-Dependenz-Grammatik des Deutschen. Technischer Bericht. Hamburg: Universität Hamburg. Internet: https://edoc.sub.uni-hamburg.de/informatik/volltexte/2014/204/ (Stand: 29.9.2020).

Franklin, Allan (1999): Can that be right? Essays on experiment, evidence, and science. (= Boston Studies in the Philosophy of Science 199). Dordrecht: Kluwer Academic.

Gilquin, Gaëtanelle (2008): Combining contrastive and interlanguage analysis to apprehend transfer: detection, explanation, evaluation. In: Gilquin, Gaëtanelle/Papp, Szilvia/Diéz-Bedmar, María Belén (Hg.): Linking up contrastive and learner corpus research. (= Language and Computers 66). Amsterdam/New York: Rodopi, S. 1–33.

Gilquin, Gaëtanelle/Paquot, Magali (2007): Spoken features in learner academic writing: identification, explanation and solution. In: Davies, Matthew/Rayson, Paul/Hunston, Susan/Danielsson, Pernilla (Hg.): Proceedings of the corpus linguistics conference. CL2007. University of Birmingham, UK, 27-30 July 2007. Birmingham: Universität Birmingham, S. 1–12.

Gilquin, Gaëtanelle/Paquot, Magali (2008): Too chatty. Learner academic writing and register variation. In: English Text Construction 1, 1, S. 41–61.

Goldberg, Adele E. (1995): Constructions: a construction grammar approach to argument structure. Chicago: University of Chicago Press.

Goldberg, Adele E. (2006): Constructions at work: The nature of generalization in language. Oxford: Oxford University Press.

Goodman, Nelson (1961): Graphs for linguistics. In: Jakobson/Roman (Hg.): Structure of language and its mathematical aspects. (= Proceedings of symposia in applied mathematics 12). Rhode Island: American Mathematical Society, S. 51–56.

Granger, Sylviane (1996): From CA to CIA and back: an integrated approach to computerized bilingual and learner corpora. In: Aijmer, Karin/Altenberg, Bengt/Johansson, Mats (Hg.): Languages in contrast: papers from a symposium on text-based cross-linguistic studies. Lund 4–5 March 1994. (= Lund studies in English 88). Lund: Lund University Press, S. 37–51.

Granger, Sylviane (1998a): The computer learner corpus: a versatile new source of data for SLA research. In: Granger, Sylviane (Hg.): Learner English on computer. London/New York: Longman, S. 3–18.
Granger, Sylviane (1998b): Prefabricated patterns in advanced EFL writing: collocations and formulae. In: Cowie, Anthony Paul (Hg.): Phraseology: theory, analysis and applications. Oxford: Clarendon Press, S. 145–160.
Granger, Sylviane (2005): Pushing back the limits of phraseology: how far can we go? In: Phraseology 2005: the many faces of phraseology. Proceedings of the phraseology 2005 conference, Université catholique de Louvain (Belgium), 13–15 October 2005. Louvain-la-Neuve: Presses universitaires de Louvain, S. 1–4.
Granger, Sylviane (2015): Contrastive interlanguage analysis: a reappraisal. In: International Journal of Learner Corpus Research 1, 1, S. 7–24.
Granger, Silviane/Gilquin, Gaëtanelle/Meunier, Fanny (Hg.) (2015): The cambridge handbook of learner corpus research. Cambridge: Cambridge University Press.
Gregory, Michelle L./Raymond, William D./Bell, Alan/Fosler-Lussier, Eric/Jurafsky, Dan (1999): The effects of collocational strength and contextual predictability in lexical production. In: Billings, Sabrina J./Boyle, John P./Griffith, Aaron M. (Hg.): Language, identity & the other. ChiPhon '99 new syntheses: multi-disciplinary approaches to basic units of speech. Theory & linguistic diversity. April 22–24, 1999. (= Chicago Linguistic Society 35). Chicago: Chicago Linguistic Society, S. 151–166.
Gries, Stefan T. (2013): 50-something years of work on collocations: What is or should be next… In: International Journal of Corpus Linguistics 18, 1, S. 137–165.
Gries, Stefan T. (2019): 15 years of collostructions: some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). In: International Journal of Corpus Linguistics 24, 3, S. 385–412.
Gries, Stefan T./Berez, Andrea L. (2017): Linguistic annotation in/for corpus linguistics. In: Ide, Nancy/Pustejovsky, James (Hg.): Handbook of linguistic annotation. Dordrecht: Springer, S. 379–409.
Gries, Stefan T./Stefanowitsch, Anatol (2004): Extending collostructional analysis: a corpus-based perspective on alternations. In: International Journal of Corpus Linguistics 9, 1, S. 97–129.
Gries, Stefan T./Wulff, Stefanie (2005): Do foreign language learners also have constructions? In: Annual Review of Cognitive Linguistics 3, 1, S. 182–200.
Haberzettl, Stefanie (2005): Der Erwerb der Verbstellungsregeln in der Zweitsprache Deutsch durch Kinder mit russischer und türkischer Muttersprache. (= Linguistische Arbeiten 495). Tübingen: Niemeyer.
Heringer, Jürgen (2001): Fehlerlexikon Deutsch als Fremdsprache. Aus Fehlern lernen: Beispiele und Diagnosen. Berlin: Cornelsen.
Hirschmann, Hagen (2011): Eine für Korpora relevante Subklassifikation adverbieller Wortarten. In: Konopka, Marek/Kubczak, Jacqueline/Mair, Christian/Štícha, František/Waßner, Ulrich Hermann (Hg.): Grammatik und Korpora 2009. Dritte Internationale Konferenz, Mannheim, 22.–24.9.2009. (= Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache 1). Tübingen: Narr, S. 157-180.
Hirschmann, Hagen (2013): Syntaktische Variation bei satzadverbialen Ausdrücken. Vortrag zur Verteidigung der Dissertation, 5.7.2013. Berlin: Humboldt-Universität zu Berlin. Vortragsfolien: https://hu.berlin/hirschmann2013dissvortr (Stand: 30.9.2020).

Hirschmann, Hagen (2015): Modifikatoren im Deutschen. Ihre Klassifizierung und varietätenspezifische Verwendung. (= Studien zur deutschen Grammatik 89). Tübingen: Stauffenburg.

Hirschmann, Hagen/Lüdeling, Anke/Rehbein, Ines/Reznicek, Marc/Zeldes, Amir (2013): Underuse of syntactic categories in Falko: a case study on modification. In: Granger, Sylviane/Gilquin, Gaëtanelle/Meunier, Fanny (Hg.): Twenty years of learner corpus research. Looking back, moving ahead. Proceedings of the first learner corpus research conference (LCR2011). (= Corpora and Language in Use Proceedings 1). Louvain-la-Neuve: Presses universitaires de Louvain, S. 223–234.

Howarth, Peter (1998): Phraseology and second language proficiency. In: Applied Linguistics 19, 1, S. 24–44.

Hyland, Ken (1998): Hedging in scientific research articles. (= Pragmatics & Beyond, New Series 54). Amsterdam/Philadelphia: Benjamins.

Jarvis, Scott (2010): Linguistic transfer. In: Jarvis, Scott/Pavlenko, Aneta (Hg.): Crosslinguistic influence in language and cognition. London/New York: Routledge, S. 61–111.

Jelinek, Fred/Bahl, Lalit/Mercer, Robert L. (1975): Design of a linguistic statistical decoder for the recognition of continuous speech. In: IEEE Transactions on Information Theory 21, 3, S. 250–256.

Jiang, Minhua (1994): Deutsche Modalpartikeln als Lehr- und Lernproblem im Fach Deutsch als Fremdsprache für Ausländer mit didaktischen Überlegungen. (= Europäische Hochschulschriften 1/Deutsche Sprache und Literatur 1478). Frankfurt a. M. (u. a.): Lang.

Kaplan, Robert B. (1966): Cultural thought patterns in inter-cultural education. In: Language learning 16, 1–2, S. 399–418.

Kilgarrif, Adam/Tugwell, David (2001): Word sketch: extraction and display of signifcant collocations for lexicography. In: Proceedings of the ACL workshop on collocations, ACL 2001, Toulouse. Stroudsburg, PA: Association for Computational Linguistics, S. 32–38.

Koplenig, Alexander (2019): Against statistical significance testing in corpus linguistics. In: Corpus Linguistics and Linguistic Theory 15, 2, S. 321–346.

Krause, Thomas (2019): ANNIS: A graph-based query system for deeply annotated text corpora. Diss. Berlin: Humboldt-Universität zu Berlin. DOI: https://doi.org/10.18452/19659.

Larsson, Tove/Paquot, Magali/Plonsky, Luke (ersch. demn.): Inter-rater reliability in learner corpus research: insights from a collaborative study on adverb placement. In: International Journal of Learner Corpus Research.

Laufer, Batia/Waldman, Tina (2011): Verb-noun collocations in second language writing: a corpus analysis of learners' English. In: Language Learning 61, 2, S. 647–672.

Lehker, Marianne/Wang, Yi (2006): Welche Unterschiede gibt es zwischen deutschen und chinesischen Aufsatzsorten aus Sicht der westlichen Textlinguistik? In: Informationen Deutsch als Fremdsprache 33, 4, S. 366–372.

Liao, Ming-Tzu/Chen, Ching-Hung (2009): Rhetorical strategies in Chinese and English: a comparison of L1 composition textbooks. In: Foreign Language Annals 42, 4, S. 695–720.

Littré, Damien (2015): Combining experimental data and corpus data: intermediate French-speaking learners and the English present. In: Corpus Linguistics and Linguistic Theory 11, 1, S. 89–126.

Liu, Lu (2005): Rhetorical education through writing instruction across cultures: a comparative analysis of select online instructional materials on argumentative writing. In: Journal of Second Language Writing 14, 1, S. 1–18.

Lorenz, Gunter (1998): Overstatement in advanced learners' writing: stylistic aspects of adjective intensification. In: Granger, Sylviane (Hg.): Learner English on computer. London/New York: Longman, S. 53–66.

Lüdeling, Anke/Hirschmann, Hagen (2015): Error annotation systems. In: Granger/Gilquin/Meunier (Hg.), S. 135–158.

Lüdeling, Anke/Hirschmann, Hagen/Shadrova, Anna (2017): Linguistic models, acquisition theories, and learner corpora: Morphological productivity in SLA research exemplified by complex verbs in German. In: Language Learning 67, S. 96–129.

Lüdeling, Anke/Doolittle, Seanna/Hirschmann, Hagen/Schmidt, Karin/Walter, Maik (2008): Das Lernerkorpus Falko. In: Deutsch als Fremdsprache 45, 2, S. 67–73.

MacWhinney, Brian (2014): Item-based patterns in early syntactic development. In: Herbst, Thomas/Schmid, Hans-Jörg/Faulhaber, Susen (Hg.): Constructions, collocations, patterns. (= Trends in Linguistics. Studies and Monographs 282). Berlin/Boston: DeGruyter, S. 33–69.

Mann, William C./Thompson, Sandra A. (1988): Rhetorical structure theory: toward a functional theory of text organization. In: Text 8, 3, S. 243–281.

Möllering, Martina (2004): The acquisition of german modal particles. A corpus-based approach. (= Linguistic Insights 10). Frankfurt a.M. (u.a.): Lang.

Nesselhauf, Nadja (2005): Collocations in a learner corpus. (= Studies in corpus linguistics 14). Amsterdam/Philadelphia: Benjamins.

Nivre, Joakim/Hall, Johan/Nilsson, Jens (2006): Maltparser: a data-driven parser-generator for dependency parsing. In: Calzolari, Nicoletta/Choukri, Khalid/Gangemi, Aldo/Maegaard, Bente/Mariani, Joseph/Odijk, Jan/Tapias, Daniel (Hg.): Proceedings of the fifth international conference on language resources and evaluation (LREC'06), May 2006, Genoa. Paris: European Language Resources Association, S. 2216–2219.

Osborne, John (2015): Transfer and learner corpus research. In: Granger/Gilquin/Meunier (Hg.), S. 333–356.

Paquot, Magali (2018): Phraseological competence: a missing component in university entrance language tests? Insights from a study of EFL learners' use of statistical collocations. In: Language Assessment Quarterly 15, 1, S. 29–43.

Paquot, Magali (2019): The phraseological dimension in interlanguage complexity research. In: Second Language Research 35, 1, S. 121–145.

Pawley, Andrew/Syder, Frances Hodgetts (1983): Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In: Richards, Jack C./Schmidt, Richard W. (Hg.): Language and communication. London/New York: Longman, S. 191–226.

Rehbein, Ines/Hirschmann, Hagen (2014a): POS tagset refinement for linguistic analysis and the impact on statistical parsing. In: Henrich, Verena/Hinrichs, Erhard/de Kok, Daniël/Osenova, Petya/Przepiórkowski, Adam (Hg.): Proceedings of the thirteenth international workshop on treebanks and linguistic theories (TLT13), December 12–13, 2014, Tübingen. Tübingen: Universität Tübingen, S. 172–183.

Rehbein, Ines/Hirschmann, Hagen (2014b): Towards a syntactically motivated analysis of modifiers in German. In: Ruppenhofer, Josef/Faaß, Gertrud (Hg.): Proceedings of the 12th edition of the KONVENS conference, Hildesheim, October 8–10, 2014. Bd. 1. Hildesheim: Universitätsverlag Hildesheim, S. 30–39.

Reiners, Ludwig (2004): Stilkunst. Ein Lehrbuch deutscher Prosa. Neubearb. v. Stefan Meyer und Jürgen Schiewe. [Originalausg. 1943]. München: Beck.

Reznicek, Marc/Walter, Maik/Schmidt, Karin/Lüdeling, Anke/Hirschmann, Hagen/Krummes, Cedric (2010): Das Falko-Handbuch: Korpusaufbau und Annotationen. Berlin: Humboldt-Universität zu Berlin.

Sag, Ivan A./Boas, Hans C./Kay, Paul (2012): Introducing sign-based construction grammar. In: Sag, Ivan A./Boas, Hans C./Kay, Paul (Hg.): Sign-based construction grammar. (= CSLI lecture notes 193). Stanford: CSLI Publications, S. 1–30.

Sauer, Simon/Lüdeling, Anke (2016): Flexible multi-layer spoken dialogue corpora. In: International Journal of Corpus Linguistics 21, 3, S. 419–438.

Shadrova, Anna (2020): Measuring coselectional constraint in learner corpora: A graph-based approach. Diss. Berlin: Humboldt-Universität zu Berlin. DOI: http://dx.doi.org/10.18452/21606.

Shadrova, Anna (i. Dr.): It may be in the structure, not the combinations: Graph metrics as an alternative to statistical measures in corpus-linguistic research. In: DhD Graph Proceedings 2019/2020.

Sinclair, John (1991): Corpus, concordance, collocation. Oxford: Oxford University Press.

Stede, Manfred (Hg.) (2016): Handbuch Textannotation: Potsdamer Kommentarkorpus 2.0. (= Potsdam Cognitive Science Series 8). Potsdam: Universitätsverlag Potsdam.

Stede, Manfred (2018): Korpusgestützte Textanalyse: Grundzüge der Ebenen-orientierten Textlinguistik. 2., überarb. Aufl. Tübingen: Narr.

Stefanowitsch, Anatol/Gries, Stefan T. (2003): Collostructions: investigating the interaction of words and constructions. In: International Journal of Corpus Linguistics 8, 2, S. 209–243.

Tardif, Twila (1996): Nouns are not always learned before verbs: evidence from Mandarin speakers' early vocabularies. Developmental Psychology 32, 3, S. 492–504.

Tardif, Twila/Shatz, Marilyn/Naigles, Letitia (1997): Caregiver speech and children's use of nouns versus verbs: a comparison of English, Italian, and Mandarin. In: Journal of Child Language 24, 3, S. 535–565.

Tsao, Feng-Fu (1983): Linguistics and written discourse in particular languages. Contrastive studies: English and Chinese (Mandarin). In: Annual Review of Applied Linguistics 3, S. 99–117.

Tomasello, Michael (2009): Constructing a language. A usage-based theory of language acquisition. Cambridge, MA: Harvard University Press.

Vyatkina, Nina A. (2007): Development of second language pragmatic competence: the data-driven teaching of German modal particles based on a learner corpus. Diss. Pennsylvania: Pennsylvania State University.

Wan, Shujun (i. Vorb.): Argumentationsstrategien von chinesischen Deutschlernern – Eine korpusbasierte Studie im Vergleich zu deutschen Muttersprachlern. Unveröff. Diss. Berlin: Humboldt-Universität zu Berlin.

Wan, Shujun/Kutschbach, Tino/Lüdeling, Aanke/Stede, Manfred (2019): RST-Tace A tool for automatic comparison and evaluation of RST trees. In: Zeldes, Amir/Das, Debopam/Galani Maziero, Erick/Desiderato Antonio, Juliano/Iruskieta, Mikel (Hg.): Proceedings of DISRPT 2019, the workshop on discourse relation parsing and treebanking, June 6, 2019, Minneapolis, MN. Stroudsburg, PA: Association for Computational Linguistics (ACL), S. 88–96.

Westpfahl, Swantje/Schmidt, Thomas/Jonietz, Jasmin/Borlinghaus, Anton (2017): STTS 2.0. Guidelines für die Annotation von POS-Tags für Transkripte gesprochener Sprache in Anlehnung an das Stuttgart Tübingen Tagset (STTS). Arbeitspapier. Mannheim: Institut für Deutsche Sprache. Internet: https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/6063 (Stand: 30.9.2020).

Wiechmann, Daniel (2008): On the computation of collostruction strength: testing measures of association as expressions of lexical bias. In: Corpus Linguistics and Linguistic Theory 4, 2, S. 253–290.

Winkler, Steffi (2011): Progressionsfolgen im DaF-Unterricht. Eine Interventionsstudie zur Vermittlung der deutschen (S)OV-Wortstellung. In: Hahn, Natalia/Roelcke, Thorsten (Hg.): Grenzen überwinden mit Deutsch: 37. Jahrestagung des Fachverbandes Deutsch als Fremdsprache an der Pädagogischen Hochschule Freiburg/Br. (= Materialien Deutsch als Fremdsprache 85). Göttingen: Universitätsverlag Göttingen, S. 193–207.

Wray, Alison (2002): Formulaic language and the lexicon. Cambridge: Cambridge University Press.

Yang, Yuchen/Yang, Zhong (2010): Problem-solution in English vs. qi-chengzhuan-he in Chinese: are they compatible discourse patterns? In: Chinese Journal of Applied Linguistics 33, 5, S. 65–79.

Zeldes, Amir (2012): Productivity in argument selection. From morphology to syntax. (= Trends in Linguistics. Studies and Monographs 260). Berlin/Boston: DeGruyter.

Zeldes, Amir (2013): Komposition als Konstruktionsnetzwerk im fortgeschrittenen L2-Deutsch. In: Zeitschrift für germanistische Linguistik 41, 2, S. 240–276.

Zeldes, Amir (2016): rstWeb – a browser-based annotation interface for rhetorical structure theory and discourse relations. In: Proceedings of NAACL-HLT 2016 (Demonstrations), San Diego, California, June 12–17, 2016. Stroudsburg, PA: Association for Computational Linguistics, S. 1–5.

Zimmermann, Klaus (1981): Warum sind die Modalpartikeln ein Lernproblem? In: Weydt, Harald (Hg.): Partikeln und Deutschunterricht. Abtönungspartikeln für Lerner des Deutschen. Heidelberg: Groos, S. 111–122.

Zinsmeister, Heije/Reznicek, Marc/Brede, Julia Ricart/Rosén, Christina/Skiba, Dirk (2012): Das Wissenschaftliche Netzwerk „Kobalt-DaF “. In: Zeitschrift für germanistische Linguistik, 40, 3, S. 457–458.