

Jahrgang 2, Nummer 2 (Dezember 2022) ISSN 2749-4411

GRAMMATIKERWERB IN DAF UND DAZ: LERNERKORPUSLINGUISTISCHE ZUGÄNGE Einleitung in die Themenausgabe

Katrin Wisniewski Otto-Friedrich-Universität Bamberg

1. Hintergrund

Dieses Themenheft der Zeitschrift KorDaF versammelt Studien, die Aspekte des Grammatikerwerbs mithilfe von Lernerproduktionen bzw. Lernerkorpora untersuchen. Der Grammatikerwerb in einer Fremd- bzw. Zweitsprache (im Folgenden: L2)¹ stellt seit dem Entstehen der L2-Erwerbsforschung vor circa 50 Jahren ein besonders intensiv beforschtes Feld dar; dabei zählt das Deutsche zu den wohl am gründlichsten erforschten L2. Eine zentrale Erkenntnisquelle stellte und stellt dabei die Analyse von gesprochenen und geschriebenen L2-Produktionen dar, vor allem seit insbesondere Selinker auf die Systematizität von Interlanguage verwiesen hatte, die stets in einem Spannungsverhältnis zur sie ebenfalls kennzeichnenden Variabilität steht (vgl. Selinker 1972). Beide Stellgrößen spielen trotz einer äußerst dynamischen theoretisch-methodischen Weiterentwicklung des Feldes nach wie vor eine zentrale Rolle bei der Untersuchung lernersprachlicher Spezifika und Entwicklungen. Unabhängig davon, was unter Grammatik verstanden wird oder welche theoretische Perspektive auf den L2-Erwerb grammatischer Phänomene eingenommen wird, rückt gerade die lernersprachliche Variabilität in jüngster Zeit wieder stark in den Fokus (vgl. Han / Tarone 2014; Gablasova 2020; Wisniewski / Czinglar / Lüdeling 2022; zur immensen Variation auch innerhalb vermeintlich homogener L1-Gruppen vgl. Shadrova et al. 2021, vgl. auch die Beiträge in diesem Band). Die Suche nach möglichst überindividuellen Erwerbssequenzen oder -stufen, Mustern in Erwerb und Gebrauch und mehr oder weniger linearen Entwicklungsprozessen, nach Zeichen von Stabilisierung oder Fossilisierung muss also immer der immensen intra- und interindividuellen Variabilität von Lernersprache Rechnung tragen, die wiederum von unterschiedlichsten Einflussfaktoren bedingt wird (z.B. Ädel 2015). Dazu zählen lernendeninterne Faktoren wie das Kontaktalter und die Kontaktintensität, motivationale Aspekte oder auch die Sprachkompetenz sowie lernendenexterne Faktoren, wo sich zuletzt beispielsweise ein maßgeblicher Einfluss von Aufgaben auf die produzierte L2 gezeigt hat (vgl. etwa Weiss 2017; Michel et al. 2019).

Lernerkorpora, die zu verstehen sind als systematisch erstellte und aufbereitete, öffentlich verfügbare digitale L2-Datensammlungen (vgl. Granger / Gilquin / Meunier 2015), bieten grundsätzlich eine ideale Ressource, um Aspekte des L2-Grammatikerwerbs empirisch zu beforschen. Gebrauchsmuster lassen sich untersuchen und besonders schwierig oder leicht zu erwerbende sprachliche Strukturen offenlegen. Je nach verfügbaren Metadaten – und diese sind für Lernerkorpora von nicht zu überschätzender Bedeutung – kann Variation lernerkorpuslinguistisch systematisch beschrieben werden. In anwendungsorientierter Perspektive wiederum lassen sich Lernerkorpora didaktisch nutzen (in der Lehre, zur Materialentwicklung), und je nach Metadaten lassen sich Bezüge auf die L2-Kompetenz der Lernenden herstellen und z.B. Fragen nach typischen sprachlichen Merkmalen auf einzelnen Niveaustufen des Gemeinsamen europäischen Referenzrahmens für Sprachen (GER) beantworten (Europarat 2001, 2020). Auch für die Diagnostik von Sprachkompetenzen spielen

¹ Unter *L2-Erwerb* wird in diesem Beitrag, so nicht anders gekennzeichnet, sowohl der überwiegend ungesteuerte Erwerb einer weiteren Sprache im zielsprachlichen Kontext (Zweitsprache) als auch der überwiegend gesteuerte Erwerb einer weiteren Sprache (Fremdsprache) verstanden). *L1* bezeichnet Sprache(n), die ab Geburt erworben werden.



Lernerkorpora eine stetig zunehmende Rolle (vgl. Götz 2022). Gegenüber anderen, nicht öffentlich verfügbaren L2-Daten bieten Lernerkorpora dabei die Vorteile von Transparenz und Nachhaltigkeit. Oftmals sind die Korpora zudem in flexiblen Mehrebenenformaten erstellt, die sich durch Nutzende erweitern lassen, beispielsweise durch zusätzliche Annotationen. Besonderes Kennzeichen der Lernerkorpusforschung ist zudem die gebräuchliche (teil-)automatisierte Annotation und Analyse. Die Möglichkeiten solcher sprachtechnologisch gestützter Analyseverfahren haben sich deutlich verbessert (vgl. Weiss et al. 2022 für eine Beispielstudie zu automatischen Analysen sprachlicher Komplexität im Unterricht und Picoral / Staples / Reppen 2021 für eine evaluierende Übersicht verschiedener Tools zum Englischen), wenn auch nicht in einem dem Englischen auch nur annähernd vergleichbaren Maße².

2. Grammatikerwerbsforschung mit Lernerkorpora: Ein kurzer Einordnungsversuch

Bei Studien zum Grammatikerwerb, die mit Korpusdaten arbeiten, lassen sich sehr grob zwei unterschiedliche Linien ausmachen, die in einem gewissen Spannungsverhältnis zueinander stehen (vgl. auch Rankin 2015; Wisniewski 2022). Eine Linie entstammt der oft stark theoriegetriebenen Zweitsprachenerwerbsforschung, die zum Ziel hat, den überwiegend ungesteuerten Erwerb einer Zweitsprache möglichst von Anbeginn zu erforschen, auch um Gemeinsamkeiten und Unterschiede zum L1-Erwerb auszumachen. Hier werden von jeher Lernerproduktionen untersucht (vgl. 2.1). Andererseits sind Lernerkorpora für die jüngere, theoretisch eklektischere Lernerkorpuslinguistik, die ihre Wurzeln in der (anglistischen) Fremdsprachendidaktik hat, die zentrale Ressource (vgl. 2.2). Datenstrukturen, Forschungsmethoden und Erkenntnisziele dieser Forschungslinien unterscheiden sich teils recht deutlich, auch wenn natürlich auch viele Gemeinsamkeiten vorherrschen und in jüngerer Zeit aus korpustechnischer Perspektive vielleicht eine langsame Konvergenz zu beobachten ist (d.h. mehr Fokus auf transparente, veröffentlichte Daten im Bereich der Zweitsprachenerwerbsforschung, mehr designbezogene Sorgfalt und umfassendere Metadaten im Bereich der Lernerkorpuslinguistik, die eher mit Daten aus dem gesteuerten Erwerb arbeitet). Im Folgenden soll dies anhand weniger Beispiele ohne jeden Anspruch auf Vollständigkeit mit Fokus auf den Grammatikerwerb des Deutschen kurz holzschnittartig illustriert werden.

2.1 Exemplarische Zugänge und Korpora der Zweitsprachenerwerbsforschung

Theoriegeleitete Zugänge zum Grammatikerwerb wurden vor allem im Bereich der Forschung zum Zweitsprachenerwerb entwickelt und verfolgt. Dabei wurden und werden die dort erstellten Korpora meist gar nicht explizit als Lernerkorpora bezeichnet (Rankin 2015 spricht von "production studies"), auch wenn mit Lernerproduktionen, teils auch mit veröffentlichten Lernerkorpora, gearbeitet wird. Solche Korpora enthalten in der Regel längsschnittliche Daten weniger Sprechender im beginnenden Zweitsprachenerwerb. Es sind – wegen des klaren Primats der Mündlichkeit im ungesteuerten

² Dennoch stoßen automatische Analysetools bei Lernersprache oft (noch) an ihre Grenzen. Automatisierte Analysen von Lernerkorpora sind voraussetzungsreich. Zielhypothesen erleichtern diese erheblich, sind aber nur für sehr wenige Lernerkorpora verfügbar, und automatisierte morphologische und syntaktische Tools (Tagger, Parser) funktionieren je nach analysiertem sprachlichem Phänomen sehr unterschiedlich genau. Lüdeling et al. (2021) kommen auch deshalb zu dem Schluss, es sei vielversprechender, manuell auf Grundlage genauer linguistischer Modelle tief annotierte, kleine bis mittlere Lernerkorpora zu analysieren, anstatt möglichst große Lernerkorpora zu konstruieren und automatisch analysieren zu wollen.



Zweitsprachenerwerb – in aller Regel gesprochene Lernerkorpora. Etliche auch ältere dieser Korpora, auf denen zentrale Studien für das Fach beruhen, sind im MPI Nijmegen im *The Language Archive* archiviert³, viele Datensammlungen wurden aber nie veröffentlicht (z.B. das *ESA-Korpus*, Nickel 2010). Es gibt nur wenige Nutzer:innenschnittstellen, die eine direkte Suche in diesen Korpora erlauben würden, etwa auch gefiltert nach bestimmten Metadaten. Einige Funktionen bietet die SLA-Talkbank⁴ mit Suchmöglichkeit auch für Daten für das Deutsche als L2⁵. Oft sind diese Korpora mit Tools der CHILDES-Gruppe aufbereitet, also nach den CHAT-Konventionen transkribiert und mit den entsprechenden Werkzeugen dann auch durchsuchbar.

Ein sehr frühes wichtiges Großprojekt war bereits in den 70er Jahren das Heidelberger Forschungsprojekt Pidgin-Deutsch mit seinem varietätengrammatischen Fokus auf lernersprachliche Entwicklungen erwachsener Arbeitsmigranten im Raum Heidelberg (zusammengefast in Klein 2021). Ein weiterer früher theoretischer Zugang entwickelte sich im ZISA-Projekt (Zweitspracherwerb italienischer und spanischer Arbeiter; vgl. Meisel / Clahsen / Pienemann 1981; Clahsen / Meisel / Pienemann 1983) und in dessen Folge unter anderem hin zur *Processability Theory* (vgl. Pienemann 1998, 2005; Lenzing / Nicholas / Roos 2019). Vereinfacht gesagt fokussiert die *Processability Theory* syntaktische und morphologische linguistische Phänomene unter der Annahme, dass diese aus Gründen der kognitiven Verarbeitbarkeit zwingend in einer hierarchischen Reihenfolge erworben werden (developmental sequences), deren Emergenz (also ihr früher systematischer Gebrauch) in Lernerproduktionen durch sogenannte Implikations- bzw. Distributionsanalysen untersucht wird. Zentral sind hier vor allem Aspekte des Verbstellungserwerbs. Die im Deutschen als Spracherwerbsstufen bezeichneten Sequenzen sind äußerst intensiv beforscht, relativ robust und haben insbesondere in DaZ-Kontexten auch zu diagnostischen Zwecken eine enorm weite Verbreitung erfahren (um nur eine Auswahl an Studien zu nennen: Diehl / Christen / Leuenberger 2000; Tschirner / Meerholz-Härle 2001; Haberzettl 2005; Jansen 2008; Czinglar 2014; Wisniewski 2020; vgl. Schlauch in diesem Band; Schwendemann in diesem Band).

In einer weiteren Forschungslinie wurde, zunächst im Rahmen des sogenannten ESF-Projekts (1982-1988, Projektleitung European Science Foundation und Max-Planck-Institut Nijmegen: Second Language Acquisition of Adult Immigrants; vgl. Perdue 1993a, 1993b), eine funktionalere Perspektive auf Lernersprache eingenommen. Auch hier standen erwachsene Zweitsprachenlernende im Fokus, die das Deutsche (bzw. andere L2) ungesteuert erwarben. Analyseschwerpunkte waren neben äußerungsstrukturellen auch diskurspragmatische Aspekte. Auch wegweisende Studien zur sogenannten basic variety entstammen diesem Projekt (vgl. Klein / Perdue 1997); Schwerpunkte waren ferner Lokalität, Temporalität und Finitheit (vgl. Perdue 1993a, 1993b). Auch dieser Forschungsstrang wurde immer wieder aufgegriffen und weiterentwickelt, bspw. in den Arbeiten rund um das konzeptorientierte p-Moll-Projekt mit dem entsprechenden Längsschnittkorpus (1985-1990, vgl. Dittmar 2012) zum Erwerb von Modalität.

Später sind nach diesen frühen wegweisenden Studien zum Beispiel auch etliche Arbeiten rund um das *DaZ-AF-Korpus*⁶ (1998-2002) entstanden, das den Deutscherwerb zweier jüngerer russischer Deutschlernender dokumentiert (vor bzw. nach der Pubertät), zum Beispiel zum Erwerb von Informationsstruktur, Finitheit und Negation, aber auch zu vielen anderen Themen (für eine Übersicht vgl. Dimroth 2008a, 2008b; vgl. auch Czinglar 2014), über die hier natürlich kein angemessener und erst recht kein vollständiger Überblick geboten werden kann. Neben diesen hier nur ganz kurz und exemplarisch angeführten Großprojekten mit ihren sehr aufwändig erstellten Korpora liegen natürlich zahllose empirische Forschungsarbeiten vor, die sich mit dem Erwerb der Grammatik des Deutschen

³ https://archive.mpi.nl/tla/islandora/object/lat%3A1839 00 0000 0000 39AD 1 (24.10.2022).

⁴ https://slabank.talkbank.org/access/ (24.10.2022).

⁵ https://sla.talkbank.org/TBB/slabank (24.10.2022).

⁶ https://hdl.handle.net/1839/00-0000-0000-0000-69D7-E (24.10.2022).



als Zweitsprache befassen und die ebenfalls mit teils großen Datenmengen arbeiten; allerdings sind diese nicht öffentlich in Form von Lernerkorpora verfügbar.

2.2 Lernerkorpuslinguistische Zugänge zum Grammatikerwerb des Deutschen: Beispiele

Die von dem oben beschriebenen Forschungszweig einigermaßen losgelöste, in den 90er Jahren aus der Fremdsprachendidaktik entstandene Lernerkorpuslinguistik versucht, den L2-Sprachgebrauch mittels möglichst großer Lernerkorpora zu beschreiben und die Erkenntnisse letztlich auch für didaktische und diagnostische Zwecke nutzbar zu machen. Sie hat nach wie vor einen sehr starken Fokus auf das Englische als L2. Eine eigene Interessensvertretung (die *Learner Corpus Association*)⁷ mit regelmäßigen Konferenzen, einem eigenen Handbuch (vgl. Granger / Gilquin / Meunier 2015), einer Zeitschrift (*International Journal of Learner Corpus Research*) und einer zunehmenden Anzahl an Publikationen⁸ verweisen auf einen vor allem international dynamischen Forschungszweig, während die Lernerkorpuslinguistik im deutschsprachigen Forschungsraum immer noch eine marginale Rolle spielt.

Zwar sind einige (kommerziellere) Lernerkorpora nur über Umwege zugänglich, dennoch gilt die öffentliche Verfügbarkeit als unumgängliches Kriterium für Lernerkorpora. Die hier gemeinten Lernerkorpora enthalten zudem typischerweise, aber keineswegs ausschließlich, kontrolliert erhobene geschriebene Daten fortgeschrittener Lernender (oftmals Essays Studierender oder in Sprachtests entstandene Texte; zu gesprochenen Lernerkorpora siehe Wisniewski 2022) und sind häufig korpuslinguistisch intensiv aufbereitet, z.B. mit verschiedenen Normalisierungs- und Annotationsebenen in Mehrebenen-Standoffverfahren (vgl. Hirschmann / Schmidt 2022). Zumindest ein Mindestmaß an Metadaten gehört zu jedem Lernerkorpus. Automatische Annotationen (z.B. von Wortarten) sind weit verbreitet.

Ein typisches und sehr bekanntes L2-Korpus des Englischen ist das *International Corpus of Learner English* (*ICLE*)⁹, zu deutschen Lernerkorpora zählen neben vielen anderen etwa die Korpora der *Falko*-Familie, *BeMaTaC*, *MERLIN* oder *DISKO*, die sämtlich über die funktionsreiche Suchund Visualisierungsschnittstelle ANNIS analysierbar sind¹⁰. An der Université de Fribourg entsteht das multimodale Schweizer Lernerkorpus (*SWIKO*, vgl. Karges / Studer / Hicks 2022), an der Europäischen Akademie Bozen sind unter anderem die KOLIPSI-Korpora¹¹ und das Korpus *Leonide* (vgl. Glaznieks et al. 2022) angesiedelt und werden über die Lernerkorpusinfrastruktur *Porta* zugänglich gemacht¹².

⁷ https://www.learnercorpusassociation.org/ (24.10.2022).

⁸ Vgl. die Bibliografie der *Learner Corpus Association*: https://www.learnercorpusassociation.org/resources/lcb/(24.10.2022).

⁹ https://uclouvain.be/en/research-institutes/ilc/cecl/icle.html (24.10.2022).

¹⁰ Falko (vgl. Reznicek et al. 2012) ist zugänglich unter https://hu-berlin.de/falko (24.10.2022). Eine Übersicht der einzelnen Korpora der Falko-Familie, die über die ANNIS-Schnittstelle zugänglich sind, findet sich unter https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko/design (24.10.2022). Informationen zu MERLIN sind unter https://merlin-platform.eu/ (24.10.2022) abrufbar; außerdem kann das Korpus über die ANNIS-Instanz der HU Berlin durchsucht werden (https://korpling.german.hu-berlin.de/annis3/, 24.10.2022) und ist hier herunterladbar (https://clarin.eurac.edu/repository/xmlui/handle/20.500.12124/6, 24.10.2022). DISKO ist ebenfalls (nach Anmeldung) über ANNIS durchsuchbar und wird im Langzeitrepositorium des Instituts für Deutsche Sprache in Mannheim zum Download verfügbar gemacht (in Vorbereitung); Informationen zum Korpus finden sich auch hier: https://home.uni-leipzig.de/sprastu/korpora/disko/, 24.10.2022). Eine internationale Übersicht zu Lernerkorpora stellt die Learner Corpus Association bereit: https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html (24.10.2022).

¹¹ Glaznieks et al. (2021): *Kolipsi-1 Corpus v1.0*, Eurac Research CLARIN Centre, http://hdl.handle.net/20.500.12124/26 sowie Glaznieks (2021): *Kolipsi-2 Corpus v1.0*, Eurac Research CLARIN Centre, http://hdl.handle.net/20.500.12124/30. 12 Ebenfalls in ANNIS und über ein Repositorium, vgl. https://www.porta.eurac.edu/ (24.10.2022).



Auch wenn die Lernerkorpuslinguistik weniger theoriegetrieben ist als die Zweitsprachenerwerbsforschung, wird die Notwendigkeit gründlicher linguistischer Modellierungen immer wieder als zentral für die Aussagekraft lernerkorpuslinguistischer Studien betont (vgl. Lüdeling / Hirschmann / Shadrova 2017; Lüdeling et al. 2021). Teils werden mittlerweile nicht mehr nur deskriptive, sondern auch explanative Ansätze des L2-Erwerbs verfolgt – denn lediglich auf der sprachlichen Oberfläche zu laborieren, ohne Spracherwerb auch erklären zu können, ist ein von Seiten der Zweitsprachenerwerbsforschung häufiger und vielleicht nicht immer ganz zu Unrecht erhobener Vorwurf gegenüber der Lernerkorpuslinguistik. Diese bedient sich vor allem gebrauchsbasierter Erwerbskonzepte (ohne strikte Trennung von Lexikon und Grammatik, z.B. Tomasello 2003; Ellis 2019), um Fragestellungen nach typischen Gebrauchsmustern und -verläufen sowie besonderen Erwerbsherausforderungen zu verfolgen. Deshalb spielen Fragen nach Standards und Gebrauchsnormen eine zentrale Rolle.

Während Fortschritt im DaZ-Kontext häufig als zeitlich determiniert operationalisiert wird, kommt im gesteuerten Grammatikerwerb (instructed second language acquisition, vgl. Loewen / Sato 2017), den die Lernerkorpuslinguistik überwiegend fokussiert, der Beschreibung sprachlicher Kompetenzen entscheidende Bedeutung zu; hierzu wird häufig auf die Niveaustufen des Gemeinsamen europäischen Referenzrahmens für Sprachen rekurriert (GER; vgl. Europarat 2001, 2020); allerdings sind nur sehr wenige Lernerkorpora auf die Niveaustufen des GER bezogen (MERLIN, DISKO).

Methodisch wird typischerweise kontrastierend gearbeitet: In aller Regel werden also unterschiedliche Gruppen miteinander verglichen (sogenannte *contrastive interlanguage analysis*, vgl. Granger et al. 2015), zum Beispiel Lernende unterschiedlicher L1 oder verschiedener Kompetenzniveaus oder aber L1- mit L2-Daten, und dann etwa hinsichtlich eines Über- oder Untergebrauchs verschiedener Strukturen analysiert. Deshalb, also weil Lernende zunächst unter Homogenitätsannahmen gruppiert werden, spielen Fragen des Umgangs mit Variabilität hier auch eine besonders wichtige Rolle. In der Lernerkorpuslinguistik wird zudem in aller Regel und in deutlichem Unterschied zur oben beschriebenen zweitsprachenerwerbstheoretischen Analyse lernersprachlicher Produktionen mit quantitativen Analyseverfahren gearbeitet. Im Zuge eines derzeit v.a. international beobachtbaren intensiven methodischen Wandels werden zunehmend komplexe statistische Verfahren verwendet und rege diskutiert (vgl. Paquot / Plonsky 2017; Gries / Deshors 2020; Gries 2021; Wulff / Gries 2021).

Grammatikbezogene Themen des Deutscherwerbs, die in Studien zu öffentlich verfügbaren deutschen Lernerkorpora untersucht wurden, sind unter etlichen anderen – und auch hier können nur einige wenige Beispiele benannt werden – beispielsweise die Attribution (vgl. Hirschmann et al. 2013; Vyatkina / Hirschmann / Golcher 2015) und allgemeiner Fragen der (syntaktischen, morphologischen ...) Komplexität, die auch international gerade äußerst intensiv beforscht werden. Bordag / Sieradz (2012) untersuchen die Verwendung von Perfekt und Passiv in *Falko*, Shadrova (2020) analysiert in einem innovativen graphenbasierten Ansatz in ihrer Dissertation Koselektionsaspekte basierend auf KobaltDaF; Zeldes (2013, 2018) analysiert Komposition und Produktivität aus konstruktionsgrammatischer Warte, Lüdeling / Hirschmann / Shadrova (2017) fokussieren komplexe Verben und ihre Produktivität. Wisniewski (2020) versucht einen Zusammenhang der morphosyntaktischen Erwerbsstufen auf die GER-Niveaus in *MERLIN* herzustellen, Wisniewski (2022) untersucht die Verwendung von *es* in verschiedenen Funktionen in den Korpora *MERLIN* und *DISKO* und bezieht dies erneut auf GER-Niveaustufen. Zudem werden Lernerkorpora auch aus computerlinguistischer Perspektive intensiv beforscht, wobei ein Fokus u.a. auf automatisierten Komplexitätsanalysen liegt (vgl. z.B. Weiss 2017; Weiss et al. 2022).



2.3 Zur nötigen Diversifizierung der Lernerkorpuslandschaft des Deutschen

Im vorliegenden Rahmen ist es unmöglich und war deshalb auch nicht Ziel, einen umfassenden und auch nur ansatzweise vollständigen Überblick über Studien zusammenzutragen, die mithilfe von Lernerkorpora den Grammatikerwerb des Deutschen genauer beleuchten. Die vorangegangen kurzen Betrachtungen dienten vielmehr einer groben Einordnung und benannten dazu einige wenige exemplarische Korpora und Studien. Dennoch drängt sich die Frage nach dem Stellenwert der immer noch jungen Lernerkorpusforschung im Fach DaF/DaZ bzw. der Germanistik auf.

Einerseits ist hier viel Positives zu vermerken. In den letzten Jahren hat sich zum einen die Anzahl und Vielfalt von öffentlichen Lernerkorpora des Deutschen stetig vergrößert; auch ältere Ressourcen sind zudem dank der Initiative des MPI Nijmegen digitalisiert und der Forschungsgemeinschaft zugänglich gemacht worden. Auch wenn deutsche Lernerkorpora u.a. hinsichtlich ihrer Größe, aber auch ihrer Diversität Lernerkorpora mit L2 Englisch nach wie vor hinterherhinken, ermöglicht das vorliegende Spektrum geschriebener, gesprochener und multimodaler Korpora doch mittlerweile umfassende systematische korpuslinguistische Analysen ganz unterschiedlicher lernersprachlicher Phänomene, wie auch eine zunehmende Zahl entsprechender Studien beweist. Außerdem ist festzuhalten: Auch wenn sich die Forschungsanliegen von Studien aus dem Kernbereich der Zweitsprachenerwerbsforschung von typischen Analysen mit Fokus auf den gesteuerten Fremdsprachenbereich natürlich ein Stück weit unterscheiden müssen, gibt es doch auch (nicht nur angesichts komplexer Erwerbs- und Mehrsprachigkeitskonstellationen, die sich kaum dem ein oder anderen Typ zuordnen lassen) viele Gemeinsamkeiten und Schnittstellen. Insbesondere bezüglich korpuslinguistischer Methoden (des Korpusdesigns, der Korpusaufarbeitung, des Forschungsdatenmanagements im Allgemeinen) scheint sich das, was oben eher zu Illustrationszwecken wie ein Graben dargestellt wurde, in Windeseile zu schließen. Die Notwendigkeit, Lernendenproduktionen nachhaltig und transparent aufzubereiten und der Forschungsgemeinschaft zur Verfügung zu stellen, ist zweifellos im Gesamtfach DaF/DaZ erkannt worden – auch wenn es bei der Umsetzung noch sehr viele, insbesondere urheber- und datenschutzrechtliche sowie technische Herausforderungen zu meistern gibt.

Andererseits liegen eben doch noch erhebliche Leerstellen in der Lernerkorpuslandschaft vor, die dringend einer Ausdifferenzierung bedürften. Es ließe sich diesbezüglich eine ganze Reihe an Desiderata benennen (vgl. dazu aber bereits Wisniewski 2022). Ganz besonders auffällig und an dieser Stelle hervorzuheben ist jedoch, dass bestehende Lernerkorpora die faktische Heterogenität der Deutschlernenden nur bruchstückhaft erfassen. Der Löwenanteil der Forschung befasst sich mit dem Zweitspracherwerb erwachsener Arbeitsmigrant:innen ohne jeden Deutschunterricht und dem weit fortgeschrittenen Deutscherwerb von Studierenden (oft an deutschen Hochschulen). Betrachtet man jedoch die gesellschaftlichen Entwicklungen der letzten Jahre, fällt auf, dass damit viele Deutschlernende gar nicht korpuslinguistisch 'erfasst' sind. Dazu zählen etwa die zahlreichen Seiteneinsteiger:innen (vgl. Schlauch in diesem Band), überhaupt jüngere Lernende (vgl. die Beiträge von Schlauch, Gamper und Selmani in diesem Band), vulnerable Zielgruppen wie bspw. (unbegleitete minderjährige) Geflüchtete (vgl. Schwendemann in diesem Band), Lernende in Integrationskursen, Berufssprachkursen, Lernende mit L1 Arabisch, Chinesisch – und nun auch Ukrainisch. Der Tatsache, dass das Deutsche in Deutschland von den unterschiedlichsten Zielgruppen auch entscheidend innerhalb von Bildungsinstitutionen und damit auf mehr oder weniger hybride Art (also sowohl gesteuert als auch ungesteuert) erworben wird, trägt die Lernerkorpusforschung weder ressourcenseitig noch hinsichtlich der Themenstellungen (z.B. Zweitschriftlernen; Erwerb von Bildungssprache; Rolle des sprachlichen (unterrichtlichen) Inputs usw.) schon ausreichend Rechnung. Gerade die Erhebung solcher Daten (also von Minderjährigen, in Bildungskontexten) trifft jedoch auf immense urheberrechtliche und datenschutzrechtliche, teils auch ethische Hürden, und ein publizierbares Korpus zu erstellen, kann für Kolleg:innen hier große Schwierigkeiten bergen.



2. Zu den Beiträgen des Themenhefts

Alle Beiträge des vorliegenden Themenhefts greifen mitten in die oben genannten Lücken hinein: Keiner der Artikel arbeitet mit einem bereits bestehenden öffentlichen deutschen Lernerkorpus; alle befassen sich mit Lernendengruppen, die in der Lernerkorpuslandschaft wenig bzw. gar nicht repräsentiert werden. In allen Studien spielen ferner die bildungsinstitutionellen Lernkontexte eine mehr oder weniger entscheidende Rolle:

Matthias Schwendemann (Leipzig/Bamberg) präsentiert eine Sonderauswertung im Kontext seiner Dissertation (vgl. Schwendemann 2022), bei der es sich um eine Studie zum gesteuerten Zweitsprachenerwerb erwachsener Deutschlernender handelt und die in einem größeren Kooperationsprojekt des Max-Planck-Instituts für Kognitions- und Neurowissenschaften und des Herder-Instituts der Universität Leipzig angesiedelt war. Das Projekt untersuchte mehrmethodisch und interdisziplinär den Deutscherwerb geflüchteter arabischer L1-Sprecher; ein nicht öffentliches, umfassendes multimodales Korpus wurde angelegt, auf dem auch die vorgelegte Studie beruht. Hier nimmt Schwendemann die Daten zweier Geschwister, die innerhalb des genannten Projekts denselben Deutschkurs besuchten, unter die Lupe. Dabei werden geschriebene, im Längsschnitt produzierte Daten hinsichtlich der durchschnittlichen Länge minimaler syntaktischer Einheiten und der erreichten Spracherwerbsstufen analysiert. Um die Variabilität der in diesem Sinne gefassten Syntaxentwicklung zu illustrieren, bedient sich der Autor innovativer Verfahren, und zwar min-max-Diagramme und sogenannter Change-Point-Analysen. Beide Methoden werden vor allem im Kontext der Theorie komplexer dynamischer Systeme (vgl. z.B. Lowie / Verspoor 2015) genutzt, bei der Variabilität als Motor für Spracherwerb betrachtet und insgesamt die Nichtlinearität von Spracherwerbsprozessen betont wird. Im Licht seiner methodologisch hochrelevanten Befunde kritisiert der Autor die Aussagekraft der weit verbreiteten Spracherwerbsstufen (als variabilitätsverdeckend), bezweifelt die Haltbarkeit ihres implikationellen Charakters und fordert eine Ausdifferenzierung der Stufen.

Lirim Selmani (Münster) stellt für seine Studie zum Erwerb von Lokalpräpositionenen in der Zweitsprache Deutsch Lernende mit albanischer L1 in den Fokus, eine bislang wenig untersuchte Lernendengruppe. Sein Korpus erhebt er v.a. unter Grundschüler:innen der dritten und vierten Klassen, die herkunftssprachlichen Unterricht erhalten und im Deutschen erstalphabetisiert wurden (N = 19). Anhand einer von den Kindern verschriftlichten Bildergeschichte analysiert Selmani dann die präpositionale lokale Relationierung vor allem hinsichtlich möglicher Einflüsse aus der L1 Albanisch, die ein deutlich weniger ausdifferenziertes Raumkategorisierungssystem besitzt als das Deutsche. Die Untersuchung macht deutlich, dass die Lernenden Schwierigkeiten mit dem Gebrauch von Lokalpräpositionen haben, und zwar in klarer Abhängigkeit von Charakteristika des jeweiligen Bezugsausdrucks; dem Autor zufolge liegen L1-Einflüsse nahe. Diese zeigten sich zudem beispielsweise in einer Übergeneralisierung der Präposition IN sowie der seltenen und nicht zielsprachenkonformen Verwendung der Präposition AN. Lirim Selmanis Beitrag verdeutlicht ferner den Bedarf weiterer, L1-spezifischer Lernerkorpora.

Jana Gamper (Gießen) befasst sich mit dem Ausbau nominaler Strukturen (genauer: Ausbau von Nominalphrasen und Nominalisierungen) als Kennzeichen des formellen Registers in Texten, die von Schüler:innen einer integrierten Gesamtschule in den Jahrgangsstufen 7 bis 11 verfasst wurden (N=107). Damit liegt der Fokus auch hier auf einer im Vergleich zu Gymnasiast:innen deutlich weniger intensiv beforschten Zielgruppe, und auch hier beruht die Analyse auf keinem bislang veröffentlichten Korpus. Zwar liegen nur eingeschränkte Metadaten vor (aus Datenschutzgründen); bei der betreffenden Schule, einem "hochheterogenen Lernort", handelt es sich aber um eine Bildungsinstitution mit 90% Schüler:innen, deren Herkunftssprache nicht das Deutsche ist. Im Ergebnis zeigt sich, dass obwohl insbesondere in Jahrgangsstufe 9 recht deutliche Entwicklungen im Sinne einer



Zunahme beider Phänomengruppen zu sehen sind, insbesondere die Heterogenität des Sprachgebrauchs der untersuchten Schüler:innen frappierend ist.

Julia Schlauchs (Gießen) Fallstudie widmet sich dem Syntaxerwerb von drei Seiteneinsteiger:innen in hessischen Intensivklassen, also der Phase (überwiegend) getrennter Beschulung vor dem Übergang in den Regelunterricht. Julia Schlauch gewährt damit einen Einblick in das im Aufbau befindliche longitudinale und multimodale Korpus SeiKo (Seiteneinsteiger:innen-Korpus), das Daten von N=15 Seiteneinsteiger:innen beinhalten und der Forschungsgemeinschaft öffentlich zur Verfügung gestellt werden soll. Die Autorin betont in ihrem Beitrag zunächst u.a. die Tatsache, dass von diesen Lernenden zwar einerseits ein zügiger Deutscherwerb erwartet wird, jedoch sowohl curriculare Grundlagen als auch Entscheidungsgrundlagen und diagnostische Tools für Übergangsentscheidungen fehlen. Auch deshalb dient hier die Verbstellung im Sinne der umfassend untersuchten Erwerbsstufen (s.o.) als Untersuchungskriterium. Ihre Analyse fokussiert v.a. die Geschwindigkeit, in der die Lernenden die unterschiedlichen Erwerbsstufen erreichen, und bezieht diese auf etliche Vorgängerstudien zu anderen Lernendengruppen. Auffallend ist vor allem, dass sich das Erwerbstempo der untersuchten drei Seiteneinsteiger:innen massiv unterscheidet.

Literatur und Ressourcen

Ädel, Annelie (2015): Variability in Learner Corpora. In: Granger, Sylviane / Gilquin, Gaetanelle / Meunier, Fanny (Hrsg.): *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 401-422.

Bordag, Denisa / Sieradz, Magdalena (2012): Erwerb von Perfekt und Passiv bei DaF-Lernern. In: *GFL-Jour-nal* 1, 1-27.

Clahsen, Harald / Meisel, Jürgen M. / Pienemann, Manfred (1983): *Deutsch als Zweitsprache. Der Spracherwerb ausländischer Arbeiter*. Tübingen: Narr.

Czinglar, Christine (2014): *Grammatikerwerb vor und nach der Pubertät. Eine Fallstudie zur Verbstellung im Deutschen als Zweitsprac*he. Berlin: De Gruyter.

Diehl, Erika / Christen, Helen / Leuenberger, Sandra (2000): *Grammatikunterricht: Alles für der Katz? Untersuchungen zum Zweitsprachenerwerb Deutsch.* Niemeyer, Tübingen.

Dimroth, Christine (2008a): Kleine Unterschiede in den Lernvoraussetzungen beim ungesteuerten Zweitspracherwerb: Welche Bereiche der Zielsprache Deutsch sind besonders betroffen? In: Ahrenholz, Bernt (Hrsg.): Zweitspracherwerb. Diagnosen, Verläufe, Voraussetzungen. Freiburg: Fillibach, 117-134.

Dimroth, Christine (2008b): Perspectives on second language acquisition at different ages. In: Philp, Jenefer / Oliver, Rhonda / Mackey, Alison (Hrsg.): Second language acquisition and the younger learner: child's play? Amsterdam: Benjamins, 53-79.

Dittmar, Norbert (2012): Das Projekt P-MoLL. Die Erlernung modaler Konzepte des Deutschen als Zweitsprache: Eine gattungsdifferenzierende und mehrebenenspezifische Langsschnittstudie. In: Ahrenholz, Bernt (Hrsg.): Einblicke in die Zweitspracherwerbsforschung und ihre methodischen Verfahren. Berlin u.a.: De Gruyter, 99-123.

Ellis, Nick C. (2019): Essentials of a Theory of Language Cognition. In: *The Modern Language Journal* 103: S1, 39-60.

Europarat (Hrsg.) (2001): Gemeinsamer Europäischer Referenzrahmen für Sprachen. Lernen, Lehren, Beurteilen. Berlin: Langenscheidt.



Europarat (Hrsg.) (2020): Gemeinsamer europäischer Referenzrahmen für Sprachen. Begleitband. Stuttgart: Klett.

Gablasova, Dana (2020): Variability. In: Tracy-Ventura, Nicole / Paquot, Magali (Hrsg.) (2020): *The Routledge Handbook of Second Language Acquisition and Corpora*. New York: Routledge, 358-369.

Glaznieks, Aivars / Frey, Jennifer-Carmen / Stopfner, Maria / Zanasi, Lorenzo / Nicolas, Lionel (2022): Leonide: A longitudinal trilingual corpus of young learners of Italian, German and English. In: *International Journal of Learner Corpus Research* 8: 1, 97-120.

Götz, Sandra (2022): Learner Corpora to Inform Testing and Assessment. In: Jablonkai, Reka R. / Csomay, Eniko (Hrsg.): *Routledge Handbook of Corpora and Language Teaching and Learning*. New York, London: Routledge, 311-325.

Granger, Sylviane / Gilquin, Gaetanelle / Meunier, Fanny (Hrsg.) (2015): *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press.

Gries, Stefan Th. / Deshors, Sandra C. (2020): Statistical Analyses of Learner Corpus Data. In: Tracy-Ventura, Nicole / Paquot, Magali (Hrsg.) (2020): *The Routledge Handbook of Second Language Acquisition and Corpora*. New York: Routledge, 119-132.

Gries, Stefan Th. (2021): (Generalized Linear) Mixed-Effects Modeling: A Learner Corpus Example. In: *Language Learning* 71: 3, 757-798.

Haberzettl, Stefanie (2005): Der Erwerb der Verbstellungsregeln in der Zweitsprache Deutsch durch Kinder mit russischer und türkischer Muttersprache. Tübingen: Narr.

Han, Zhao Hong / Tarone, Elaine (Hrsg.) (2014): *Interlanguage: Forty Years Later*. Amsterdam: John Benjamins.

Hirschmann, Hagen / Schmidt, Thomas (2022): Gesprochene Lernerkorpora: Methodisch-technische Aspekte der Erhebung, Erschließung und Nutzung. In: *Zeitschrift für germanistische Linguistik* 50: 1, 36-81.

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, Gaetanelle / Menuier, Fanny (Hrsg.): *Twenty Years of Learner Corpus Research. Looking Back, Moving Ahead.* Louvain: Presses Universitaires de Louvain, 223-234.

Jansen, Louise (2008): Acquisition of German word order in tutored learners: A cross-sectional study in a wider theoretical context. In: *Language Learning* 58: 1, 185-231.

Karges, Katharina / Studer, Thomas / Hicks, Nina S. (2022): Lernersprache, Aufgabe und Modalität: Beobachtungen zu Texten aus dem Schweizer Lernerkorpus SWIKO. In: *Zeitschrift für germanistische Linguistik* 50: 1, 104-130.

Klein, Wolfgang (2021): Das "Heidelberger Forschungsprojekt Pidgin-Deutsch" und die Folgen. In: Ahrenholz, Bernt / Rost-Roth, Martina (Hrsg.): Ein Blick zurück nach vorn. Frühe deutsche Forschung zu Zweitspracherwerb, Migration, Mehrsprachigkeit und zweitsprachbezogener Sprachdidaktik sowie ihre Bedeutung heute. Berlin u.a.: De Gruyter, 51-96.

Klein, Wolfgang / Perdue, Clive (1997): The Basic Variety (or: Couldn't natural languages be much simpler?). In: *Second Language Research* 13: 4, 301-347.

Lenzing, Anke / Nicholas, Howard / Roos, Jana (2019): Widening contexts for Processability Theory: Theories and issues. Amsterdam: John Benjamins.

Loewen, Shawn / Sato, Masatoshi (2017): *The Routledge handbook of instructed second language acquisition*. New York: Routledge.



Lowie, Wander / Verspoor, Marjolijn (2015): Variability and Variation in Second Language Acquisition Orders: A Dynamic Reevaluation. *Language Learning* 65: 1, 63-88.

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: S1, 96-129.

Lüdeling, Anke / Hirschmann, Hagen / Shadrova, Anna / Wan, Shujun (2021): Tiefe Analyse von Lernerkorpora. In: Lobin, Henning / Witt, Andreas / Wöllstein, Angelika (Hrsg.): *Deutsch in Europa. Sprachpolitisch, grammatisch, methodisch.* Berlin u.a.: De Gruyter, 235-283.

Meisel, Jürgen M. / Clahsen, Harald / Pienemann, Manfred (1981): On determining developmental stages in natural second language acquisition. In: *Studies in Second Language Acquisition* 3: 2, 109-135.

Michel, Marije / Murakami, Akira / Alexopoulou, Theodora / Meurers, Detmar (2019): Effects of Task Type on Morphosyntactic Complexity across Proficiency: Evidence from a Large Learner Corpus of A1 to C2 Writings. In: *Instructed Second Language Acquisition* 3: 2, 124-152.

Nickel, Aneta (2010): Sprachliche Entwicklung am Beispiel von Präpositionalphrasen. Morphologische Analysen am ESA-Korpus. In: Rost-Roth, Martina (Hrsg.): *DaZ-Spracherwerb und Sprachförderung Deutsch als Zweitsprache*. Freiburg i. Br.: Fillibach, 53-77.

Paquot, Magali / Plonsky, Luke (2017): Quantitative research methods and study quality in learner corpus research. In: *International Journal of Learner Corpus Research* 3: 1, 61-94.

Perdue, Clive (Hrsg.) (1993a): Adult Language Acquisition. Vol. 1: Field Methods. Cambridge: Cambridge University Press.

Perdue, Clive (Hrsg.) (1993b): Adult Language Acquisition. Vol. 2: The Results. Cambridge: Cambridge University Press.

Picoral, Adriana / Staples, Shelley / Reppen, Randi (2021): Automated annotation of learner English. In: *International Journal of Learner Corpus Research* 7: 1, 17-52.

Pienemann, Manfred (1998): Language processing and second language development. Processability theory. Amsterdam: John Benjamins.

Pienemann, Manfred (Hrsg.). (2005). Cross-linguistic aspects of processability theory. Amsterdam: John Benjamins.

Rankin, Tom (2015). Learner corpora and grammar. In: Granger, Sylviane / Gilquin, Gaetanelle / Meunier, Fanny (Hrsg.) (2015): *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press, 231-254.

Reznicek, Mark / Lüdeling, Anke / Krummes, Cedric / Schwantuschke, Franziska / Walter, Maik / Schmidt, Karin / Hirschmann, Hagen / Andreas, Thorsten (2012): *Das Falko-Handbuch: Korpusaufbau und Annotationen.* Berlin: Institut für deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin. https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko/standardseite (24.10.2022).

Schwendemann, Matthias (2022): Die Entwicklung syntaktischer Strukturen. Eine Längsschnittstudie anhand schriftlicher Sprachdaten erwachsener Deutschlernender mit der Erstsprache Arabisch. Leipzig. Eingereichte Dissertation.

Selinker, Larry (1972): Interlanguage. In: IRAL - International Review of Applied Linguistics in Language Teaching 10, 1-4.

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



Shadrova, Anna / Linscheid, Pia / Lukassek, Julia / Lüdeling, Anke / Schneider, Sarah (2021): A Challenge for Contrastive L1/L2 Corpus Studies: Large Inter- and Intra-Individual Variation Across Morphological, but Not Global Syntactic Categories in Task-Based Corpus Data of a Homogeneous L1 German Group. In: *Frontiers in Psychology* 12, 52-67.

Tomasello, Michael (2003): Constructing a language. Cambridge, MA: Harvard University Press.

Tschirner, Erwin / Meerholz-Härle, Birgit (2001): Processability Theory: Eine empirische Untersuchung. In: Aguado, Karin / Riemer, Claudia (Hrsg.): Wege und Ziele: Zur Theorie, Empirie und Praxis des Deutschen als Fremdsprache (und anderer Fremdsprachen). Festschrift für Gert Henrici. Hohengehren: Schneider-Verlag, 155-175.

Vyatkina, Nina / Hirschmann, Hagen / Golcher, Felix (2015): Syntactic modification at early stages of L2 German writing development: A longitudinal learner corpus study. In: *Journal of Second Language Writing* 29, 28-50.

Weiss, Zarah / Lange-Schubert, Kim / Geist, Barbara / Meurers, Detmar (2022): Sprachliche Komplexität im Unterricht. In: *Zeitschrift für germanistische Linguistik* 50: 1, 159-201.

Weiss, Zarah (2017): Using Measures of Linguistic Complexity to Assess German L2 Proficiency in Learner Corpora under Consideration of Task-Effects. MA thesis in computational linguistics. Tübingen: Eberhard-Karls-Universität Tübingen. http://www.sfs.uni-tuebingen.de/~zweiss/ (24.10.2022).

Wisniewski, Katrin / Czinglar, Christine / Lüdeling, Anke (2022): Zum Umgang mit Variation in der Lernersprachenanalyse. Perspektiven aus und für DaF/DaZ. In: *Deutsch als Fremdsprache* 4, 195-206.

Wisniewski, Katrin (2020): SLA developmental stages in a CEFR-related learner corpus: Inversion and verbend structures in German L2. In: *International Journal of Learner Corpus Linguistics* 6: 1, 1-37.

Wisniewski, Katrin (2022): Gesprochene Lernerkorpora des Deutschen: Eine Bestandsaufnahme. In: Zeitschrift für germanistische Linguistik 50: 1, 1-35.

Wulff, Stefanie / Gries, Stefan Th. (2021): Exploring Individual Variation in Learner Corpus Research: Methodological Suggestions. In: Le Bruyn, Bert / Paquot, Magali (Hrsg.): *Learner Corpus Research Meets Second Language Acquisition*. Cambridge: Cambridge University Press, 191-213.

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

Zeldes, Amir (2018): Compounds and productivity in advanced L2 German writing. A constructional approach. In: Tyler, Andrea A. / Ortega, Lourdes / Uno, Mariko / Park, Hae In (Hrsg.): *Usage-inspired L2 Instruction: Researched pedagogy*. Amsterdam: John Benjamins, 237-265.



Biographische Notiz: Katrin Wisniewski ist Professorin für Deutsche Sprachwissenschaft und Deutsch als Fremdsprache an der Otto-Friedrich-Universität in Bamberg. Neben Schwerpunkten in der Sprachdiagnostik und der L2-Erwerbsforschung befasst sie sich mit Lernerkorpora (*MERLIN*, *DISKO*). Derzeit arbeitet sie im BMBF-geförderten Projekt DAKODA mit an der Erstellung einer korpusübergreifenden Plattform zur Analyse von Variabilität innerhalb von Erwerbsstufen des Deutschen.

Kontaktanschrift:

Prof. Dr. Katrin Wisniewski Otto-Friedrich-Universität Bamberg Institut für Germanistik Hornthalstraße 2 94067 Bamberg katrin.wisniewski@uni-bamberg.de

