Studien Deutsch als Fremd- und Zweitsprache

Matthias Schwendemann

Die Entwicklung syntaktischer Strukturen

Eine Längsschnittstudie anhand schriftlicher Sprachdaten erwachsener Deutschlernender mit der Erstsprache Arabisch

ESV ERICH SCHMIDT VERLAG

Open Access
frei verfügbar

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Studien Deutsch als Fremd- und Zweitsprache

Herausgegeben von
Christian Fandrych, Marina Foschi Albert,
Karen Schramm und Maria Thurmair

Band 17

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Die Entwicklung syntaktischer Strukturen

Eine Längsschnittstudie anhand schriftlicher Sprachdaten erwachsener Deutschlernender mit der Erstsprache Arabisch

Von

Matthias Schwendemann

ERICH SCHMIDT VERLAG

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über https://dnb.dnb.de abrufbar.

Weitere Informationen zu diesem Titel finden Sie im Internet unter
ESV.info/978-3-503-21222-4

Dieses Werk ist lizenziert unter der Creative-Commons-Attribution-Non-Commercial-NoDerivates 4.0 Lizenz (BY-NC-ND).

Diese Lizenz erlaubt die private Nutzung, gestattet aber keine Bearbeitung und keine kommerzielle Nutzung.

Weitere Informationen unter *https://creativecommons.org/licenses/by-nc-nd/4.0/*

Sie finden den Anhang des Buches frei zugänglich unter:
https://anhang-syntaktische-strukturen.ESV.info

zugl. Dissertation Universität Leipzig

Umschlaggestaltung unter Verwendung
einer Fotografie von © vexworldwide, stock.adobe.com

Gedrucktes Werk: ISBN 978-3-503-21222-4
eBook: ISBN 978-3-503-21223-1
DOI https://doi.org/10.37307/b.978-3-503-21223-1

Alle Rechte vorbehalten
© Erich Schmidt Verlag GmbH & Co. KG, Berlin 2023
www.ESV.info

Satz: Arnold & Domnick, Leipzig
Druck und Bindung: Hubert & Co., Göttingen

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Danksagung

Dass diese Arbeit fertiggestellt und veröffentlicht werden konnte, verdanke ich sehr vielen Menschen. Zunächst gilt mein Dank Prof. Friederici, die durch ihre Idee und ihr Vertrauen einerseits diese Studie und andererseits die vorliegende Arbeit möglich gemacht hat.

Dann gilt meine tiefe Dankbarkeit Prof. Fandrych und Prof. Wisniewski, die mich durch die vielen guten Ratschläge und aufbauenden Worte und ihr großes Vertrauen zu jedem Zeitpunkt enorm unterstützt haben, auch und gerade dann, als ich während der Arbeit an dieser Dissertation gar nicht mehr wusste, wie es weitergeht.

Ich danke außerdem dem Forschungsteam am Max-Planck-Institut und hier vor allem Alfred, Tomás, Helyne und Xuehu. Da in dieser Studie Lernersprache analysiert wird, gilt mein ganz besonderer Dank den Lernerinnen und Lernern, deren unglaublich große Lernanstrengungen und Lernleistungen dieser Arbeit zugrunde liegen. In diesem Kontext dürfen ebenso die Lehrkräfte nicht unerwähnt bleiben, die mit den Teilnehmenden zusammengearbeitet und sie auf diesem Weg begleitet haben. Außerdem möchte ich den Studienassistentinnen und Studienassistenten danken, die unermüdlich dafür gearbeitet haben, die Kommunikation zwischen Forschenden und Teilnehmenden sicherzustellen.

Sehr dankbar bin ich außerdem den Kolleg:innen am Herder-Institut, durch die sich der Weg zur Arbeit immer auch ein wenig wie der Weg nach Hause anfühlt. Hier möchte ich mich besonders bei Annette Portmann bedanken, die mir mit wertvollen Kommentaren und Fragen sehr weitergeholfen hat. Für Kommentare und konstruktive Kritik möchte ich mich außerdem bei Ahmed Schrör bedanken, der mir gerade in der Schlussphase dieser Arbeit immer wieder dabei geholfen hat, nicht alles hinzuschmeißen und doch noch ein wenig weiter zu basteln und die eine oder andere Formulierung zu überdenken. Ein ganz besonderer Dank gilt zudem Frau Dr. Lehnen und Frau Posingies vom Erich Schmidt Verlag, die mich mit ihrem Fachwissen und ihrer Professionalität durch jeden Schritt des Veröffentlichungsprozesses geführt haben.

Zum Schluss möchte ich meiner Frau Julia, meinen Eltern und meinen Schwestern danken, die ebenfalls immer wieder viel Korrekturlesen mussten. Vor allem aber haben sie mich mit unendlicher Geduld und viel Liebe und sehr viel Verständnis immer begleitet und durch die Zeit der Arbeit an dieser Dissertation getragen. Priska, diese Arbeit ist für dich.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Inhaltsverzeichnis

Danksagung 5

1 Einleitung 11
- 1.1 Problemstellung und Relevanz 11
- 1.2 Erkenntnisinteresse und Forschungsfragen 18
- 1.3 Hybride Lern- und Erwerbssituation zwischen DaF und DaZ 20
- 1.4 Abgrenzung der vorliegenden Arbeit zur Gesamtstudie 22
- 1.5 Aufbau der Arbeit 24

2 Forschungsstand 27
- 2.1 Unterschiede zwischen Erst- und Zweitspracherwerb 27
- 2.2 Erforschung und Analyse von Lernersprache 32
- 2.3 Schreiben in Deutsch als Zweitsprache 36
- 2.4 Processability Theory und Profilanalyse 41
  - 2.4.1 Grundannahmen und theoretischer Hintergrund 41
  - 2.4.2 Empirisch-methodische Entwicklung der *Processability Theory* 51
  - 2.4.3 Profilanalyse 61
  - 2.4.4 Kritik und Grenzen erwerbsstufenbezogener Ansätze 65
  - 2.4.5 Bedeutung für DaF/DaZ 68
  - 2.4.6 Zusammenfassung 69
- 2.5 *Complex Dynamic Systems Theory* im Zweitspracherwerb 72
  - 2.5.1 Grundannahmen und theoretischer Hintergrund 74
  - 2.5.2 Empirische-methodische Entwicklung CDST-basierter Studien 83
  - 2.5.3 Kritik und Grenzen von Ansätzen mit komplex-dynamischer Perspektive 95
  - 2.5.4 Bedeutung für DaF/DaZ 96
  - 2.5.5 Zusammenfassung 98

3 Methodisches Vorgehen 101
- 3.1 Methodische Grundlagen und Datenbasis 102
  - 3.1.1 Rekrutierung der Teilnehmenden und Zugangsvoraussetzungen 103
  - 3.1.2 Teilnehmende innerhalb der Gesamtstudie 106
  - 3.1.3 Longitudinaler Aufbau und zeitlicher Ablauf der Studie 112
  - 3.1.4 Materialzugang und Messpunkte 118
  - 3.1.5 Experimentelle Manipulation 120
  - 3.1.6 Limitationen der Studie 123
  - 3.1.7 Zusammenfassung der Grundlagen 129

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

3.2 Methodisches Vorgehen: Gruppendaten ... 130
3.2.1 Teilstichprobe für Forschungsfragen 1 und 2 ... 131
3.2.2 Unterschiede zwischen Teilstichprobe und ausgeschlossenem Datenset ... 133
3.2.3 Datengrundlage für Forschungsfragen 1, 2 und 3: Sprachtests ... 135
3.2.4 Methodisches Vorgehen: Emergenz- und Profilanalyse ... 142
3.2.5 Methodisches Vorgehen: Longitudinale Clusteranalysen ... 146
3.2.6 Zusammenfassung des methodischen Vorgehens bezüglich der Gruppendaten ... 156
3.3 Methodisches Vorgehen: Einzelfallanalysen ... 156
3.3.1 Beschreibung der beiden Lernenden TN024 und TN025 ... 157
3.3.2 Auswahl der Texte ... 160
3.3.3 Methodisches Vorgehen: (Grafische) Trendanalysen ... 162
3.3.4 Methodisches Vorgehen: Generelle und gleitende Korrelationen ... 164
3.3.5 Methodisches Vorgehen: *Change Point*-Analysen ... 165
3.3.6 Zusammenfassung des methodischen Vorgehens in Bezug auf die Einzelfallanalysen ... 168

4 Ergebnisse ... 171
4.1 Ergebnisse: Gruppendaten ... 171
4.1.1 Ergebnisse: Deskriptive Statistiken zu TP1, TP2 und TP3 ... 171
4.1.2 Ergebnisse: Emergenz- und Profilanalysen ... 176
4.1.3 Ergebnisse: Longitudinale Clusteranalysen ... 190
4.1.4 Zusammenfassung der Ergebnisse auf Gruppenebene ... 213
4.2 Ergebnisse: Einzelfallanalysen ... 215
4.2.1 Deskriptive Beschreibung ... 216
4.2.2 Grafische Analyse: Variabilität ... 228
4.2.3 Grafische Trendanalyse der Interaktionen: Wortstellungsmuster ... 236
4.2.4 Grafische Trendanalyse der Interaktionen: Syntaxvariablen ... 246
4.2.5 Zwischenfazit: Grafische Interpretation ... 252
4.2.6 Generelle und gleitende Korrelationen: Wortstellungsmuster ... 254
4.2.7 Generelle und gleitende Korrelationen: Syntaxvariablen ... 265
4.2.8 *Change Point*-Analysen ... 272
4.2.9 Zusammenfassung: Einzelfallanalysen ... 282
4.3 Zusammenfassung der Ergebnisse ... 284

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

5 Diskussion und Ausblick 289
5.1 Heterogenität als Ausgangspunkt 289
5.2 Methodische Diskussion und Limitationen der Studie 292
5.3 Variabilität und Erwerbsstufen 299
5.4 Experimentelle Manipulation 302
5.5 Ausblick 304

6 Literaturverzeichnis 307

7 Tabellenverzeichnis 337

8 Abbildungsverzeichnis 340

9 Abkürzungsverzeichnis 343

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

1 Einleitung

1.1 Problemstellung und Relevanz

Fremd- und zweitsprachliche Erwerbs- bzw. Entwicklungsprozesse werden im deutschsprachigen Fachzusammenhang Deutsch als Fremdsprache / Deutsch als Zweitsprache (DaF/DaZ) seit Jahrzehnten in verschiedenen Kontexten systematisch untersucht (vgl. für einen historischen Abschnitt Ahrenholz 2020b). Bis heute bleiben jedoch Fragen über den Ablauf des Zweitspracherwerbs und seiner Unterschiede zum Erstspracherwerb offen oder werden neu gestellt.[1] Einen zentralen Ausgangspunkt der Zweitspracherwerbsforschung bildet das Interesse daran, den Erwerbsstand grammatischen Wissens der Lernenden zu bestimmten Zeitpunkten während des Erwerbprozesses zu bestimmen. In diesem Zusammenhang soll unter anderem die Frage beantwortet werden, in welcher Weise und Reihenfolge Strukturen erworben werden und welche sprachlichen Funktionen die Lernenden jeweils durch den Erwerb der betroffenen Strukturen realisieren können (vgl. Fandrych 2013). Aus Sicht der Didaktik des Deutschen als Zweit- und Fremdsprache sind diese grundlegenden Fragen eng mit Evaluierung und Weiterentwicklung bestehender Unterrichtsmethoden verbunden.

Viele der stark rezipierten Studien im Kontext DaF/DaZ waren als Einzelfallanalysen konzipiert, die eine geringe Anzahl von Lernerinnen und Lernern sehr genau in ihren Entwicklungen in den Blick nahmen (siehe unten; vgl. Kapitel 2). Dies führte zwar einerseits zu einer sehr detaillierten Modellierung individueller Lernverläufe (vgl. etwa Pienemann 1998; vgl. Czinglar 2014 u. a.), erschwerte aber andererseits mögliche Validierungen der herausgearbeiteten Ergebnisse über die untersuchten Einzelfälle hinaus. Zweitspracherwerbsstudien wurden außerdem oft im vorschulischen oder frühschulischen Bereich durchgeführt. Über erwachsene Lernende existieren hingegen deutlich weniger Daten (vgl. Fandrych 2013:167).

Als in den letzten Jahren viele Lernende mit relativ homogenen Sprachhintergründen auf einmal neu nach Deutschland kamen, eröffnete sich gleichzeitig eine Reihe von Forschungsperspektiven. Es bot sich beispielsweise die Möglichkeit zur Realisierung von longitudinalen Zweitspracherwerbsstudien größeren Umfangs mit erwachsenen Lernenden. Dass dies einerseits ein großes Desiderat

1 Zur begrifflichen Unterscheidung von Erst- und Zweitspracherwerb vgl. Ahrenholz (2020a). In der vorliegenden Arbeit wird der Begriff Erstsprache (L1) verwendet und der Begriff Muttersprache vermieden. Ebenso wird in dieser Arbeit in der Regel der Begriff Zweitspracherwerb bzw. Zweitsprachentwicklung gebraucht. In manchen Fällen wird von L2 gesprochen. Der englische Begriff *Second Language Acquisition* (SLA) wird weitgehend deckungsgleich gesehen. Erkenntnisse der Tertiärsprachenforschung (Deutsch als L3 oder Deutsch als L3nE; vgl. Hufeisen 2003) werden hier und im Folgenden nicht systematisch einbezogen.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

darstellte und immer noch darstellt, zeigen in jüngster Zeit Veröffentlichungen aus den deutschsprachigen Ländern, wie etwa die Studie „Deutsch im Beruf: Die sprachlich-kommunikative Integration von Flüchtlingen", die das Leibniz-Institut für Deutsche Sprache in Mannheim (IDS) und das Goethe-Institut (GI) in Kooperation durchgeführt haben und in der die sprachlich-sozialen Hintergründe von Geflüchteten im Kontrast zu anderen Zugewanderten verglichen wurden (vgl. Hünlich et al. 2018 und Cindark / Hünlich 2019). Andererseits erschienen viele sprachdidaktisch orientierte Veröffentlichungen, die die Relevanz für den Fachzusammenhang DaF/DaZ verdeutlichen und die besondere Situation von Lernenden mit Flucht- bzw. Migrationshintergrund in den Blick nahmen. Hier sind etwa das Sonderheft von Fremdsprache Deutsch „Deutschunterricht für Lernende mit Migrationshintergrund" (vgl. Fandrych et al. 2016), der Band 97 der Reihe Materialien Deutsch als Fremdsprache „Wie schaffen wir das? Beiträge zur sprachlichen Integration geflüchteter Menschen" (Middeke et al. 2017) oder die Themenschwerpunkte der Zeitschrift für Interkulturellen Fremdsprachenunterricht „Ziel berufliche Integration: DaF-Unterricht für geflüchtete Jugendliche und junge Erwachsene" (Merkelbach 2020) und der Zeitschrift Informationen Deutsch als Fremdsprache „Deutsch für Seiteneinsteiger/innen" (Gamper et al. 2020) zu nennen.

Die Grundlage der vorliegenden Untersuchung bilden vor diesem Hintergrund Daten, die zwischen 2016 und 2017 in einer Längsschnittstudie zum Zweitspracherwerb erwachsener Lernender mit Arabisch als Erstsprache am Max-Planck-Institut für Kognitions- und Neurowissenschaften (Leipzig) in Kooperation mit dem Herder-Institut der Universität Leipzig erhoben wurden. Die leitende Frage der Studie war, wie sich das Gehirn den Anforderungen einer neu zu lernenden Zweitsprache anpasst, wobei die Plastizität des menschlichen Sprachnetzwerkes und die sich verändernden neuronalen Wechselwirkungen und Verbindungen innerhalb dieses Systems die Kerninteressen der Forschung bildeten. Sowohl aus neurowissenschaftlicher als auch aus der spracherwerbsbezogenen Perspektive von DaF/DaZ spielen syntaktische und semantische Prozesse als Grundfunktionen von Sprache (vgl. Friederici 2013: 250) eine zentrale Rolle im Design der Studie und in der Formulierung der Forschungsfragen (vgl. Kapitel 1.2). Aus der Perspektive des Forschungszusammenhangs DaF/DaZ sind die vorliegende Studie und die in ihrem Rahmen gesammelten Daten aus verschiedenen Gründen aktuell und relevant. Einerseits erlaubt es der Aufbau der Studie, im Längsschnitt über insgesamt 17 Monate hinweg exakte und weitreichende Daten über den Erwerb bzw. die Entwicklung des Deutschen als Zweitsprache zu sammeln. Eine weitere Forschungsperspektive eröffnet sich durch die erstsprachlich homogene Gruppe von Lernerinnen und Lernern, die sich ausschließlich aus Erwachsenen mit der Erstsprache Arabisch zusammensetzt. Die Zielgruppe erwachsener Lernender mit L1 Arabisch wird über die nächsten Jahre für die Didaktik und die Forschung im Kontext von DaF/DaZ eine Lernendengruppe von hohem Stellenwert bleiben. Eine weitere Forschungsebene ergibt sich aus der experimentellen Manipulation innerhalb der Studie, im

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Rahmen derer die Gruppe der Lernenden in zwei verschiedene Lerngruppen eingeteilt wurde. Eine Studie dieser Größe und Dauer kann wichtige Impulse und Erkenntnisse für die Zweitsprachdidaktik im Bereich der Erwachsenenbildung liefern, die zukünftig in fundierte Diskussionen um die sinnvolle Weiterentwicklung und die empirische Überprüfung von Lehr- und Lernmaterialien und Lehr- und Lernmethoden im Kontext DaF/DaZ überführt werden könnten.

In der vorliegenden Arbeit wird durch die Analyse schriftlicher Produktionen die Entwicklung ausgewählter syntaktischer Strukturen in den lernersprachlichen Systemen erwachsener Lernender untersucht. Schriftliche Produktionen sind oft komplexer als mündliche und bilden daher unter Umständen genauer ab, was Lernende in einer L2 ‚können', da mehr Zeit zur Verfügung steht, um eigene Produktionen zu planen (vgl. Verspoor / Schmid / Xu 2012: 239). In der bisherigen Forschung wurden viele Parameter vorgeschlagen und umfangreich evaluiert, die unterschiedliche Aspekte der zweitsprachlichen Progression zu erfassen scheinen und in Lernersprachen sichtbar machen (dazu beispielhaft Lu 2011 und Vyatkina 2013; vgl. ebenso die Ausführungen in Kapitel 2 zum Forschungsstand). Einige dieser Parameter bzw. Maße der syntaktischen Entwicklung werden in dieser Arbeit zur Beantwortung der Forschungsfragen operationalisiert (zum methodischen Vorgehen vgl. Kapitel 3).

Als Lernersprache wird die sich dynamisch verändernde Repräsentation der zu erlernenden Sprache gesehen, die sich in der Auseinandersetzung der Lernerinnen und Lerner mit ebendieser entwickelt.[2] Ein wesentliches Merkmal von Lernersprachen sind ihre große Variabilität und ihr selten stabiler Zustand (vgl. Selinker 1972; vgl. Pienemann 1998; vgl. de Bot / Lowie Verspoor 2007 u.a.). Variabilität bezieht sich auf die Formen- und Strukturvielfalt in den beobachteten Lernersprachen, in denen sich sowohl zielsprachliche als auch nicht-zielsprachliche Formen und Strukturen finden können. Lernersprachen werden oft indirekt mithilfe von schriftlichen oder mündlichen Produktionsdaten untersucht. Der große Vorteil von schriftlichen Lernersprachdaten ist ihre relativ mühelose Zugänglichkeit und schnelle Erfassbarkeit. Mithilfe (halb-)automatisierter Verfahren können viele Daten in kurzer Zeit erfasst und für weitere Analysen

2 Zum Begriff Lernersprache vgl. u.a. grundlegend Selinker (1972), außerdem die Ausführungen in Ahrenholz (2014) und Marx / Mehlhorn (2016); dazu in dieser Arbeit Kapitel 2.2. Der Begriff *Interlanguage* hat sich seit dem gleichnamigen Aufsatz von Selinker weithin etabliert. Diese Sprachform ist „sowohl systematischer als auch dynamischer Natur" (Ahrenholz 2014: 167). In der vorliegenden Arbeit wird in diesem Verständnis von *Interlanguage* für von Lernenden produzierte (und rezipierte) Strukturen der Begriff Lernersprache verwendet. Da sich der Begriff Lernersprache in der deutschsprachigen Forschung weitgehend etabliert zu haben scheint, wird er in dieser Arbeit unverändert und nicht in einer gendergerechten Schreibweise benutzt. Dies bezieht sich ebenfalls auf das zugehörige Adjektiv ‚lernersprachlich'. Alle anderen sprachlichen Formen werden stets in gendergerechter Sprache wiedergegeben.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

aufbereitet werden, was schriftliche Daten zu einer häufig operationalisierten Datengrundlage in unterschiedlichen Studien zum Zweitspracherwerb bzw. zur Zweitsprachentwicklung macht (vgl. Fogal / Verspoor 2020: X).

Die Anfänge der Erforschung von zweitsprachlichen Erwerbs- und Entwicklungsprozessen gehen in Deutschland auf die Zeit des großen Wirtschaftswachstums und des daraus resultierenden Fachkräftemangel in den 1950er-Jahren zurück (vgl. Bryant / Rinker 2021: 286f.), in dessen Folge eine große Anzahl von Fachkräften ohne Deutschkenntnisse angeworben wurde. Die ersten Forschungsarbeiten, die in diesem Kontext entstehen, finden sich in den späten 1960er-Jahren und stellen exemplarische und eher defizitorientierte Untersuchungen der Sprache von Gastarbeiterinnen und Gastarbeitern dar. Clyne (1968) geht noch davon aus, dass die Gastarbeiter/-innen seiner Studie Sprache in Alltagssituationen nutzen, um „primitivste […] Notwendigkeiten" (Clyne 1968: 130) auszudrücken. Trotz seiner insgesamt defizitorientierten Analyse von Lernersprache stellt Clyne heraus, dass die zum Teil bruchstückhaften Äußerungen der untersuchten Lernenden funktional ähnlich effektiv sind wie eine ‚korrekt' ausformulierte Äußerung (vgl. ebd.: 136f.). Mehrere große Forschungsprojekte nahmen daraufhin den Forschungsgegenstand in den 1970er-Jahren auf.[3] Ahrenholz (2020b) nennt „das Kieler Projekt (Wode 1993), dann das *Heidelberger Forschungsprojekt Pidgin Deutsch* (HDP 1977), Ende der 1970er-Jahre das *ZISA*-Projekt (Zweitspracherwerb italienischer (portugiesischer) und spanischer Arbeiter; Clahsen / Meisel / Pienemann 1983); […]" (Ahrenholz 2020b: 107; [Hervorhebungen im Original; MS]). Darauf folgen wenig später in den 1980er-Jahren die Projekte P-MoLL (Dittmar / Reich / Schumacher / Skiba / Terborg 1990) und *Second Language Acquisition by Adult Immigrants* (Perdue 1993; vgl. Ahrenholz 2020b: 107).[4]

3 Die folgenden überblicksartigen Ausführungen folgen im Wesentlichen Ahrenholz (2020b). Andere sehr umfangreiche Darstellungen dieser frühen Zweitspracherwerbsstudien finden sich in Ahrenholz 2012a; ein englischsprachiger Überblick über frühe Theoriebildung im Bereich der Zweitspracherwerbsforschung findet sich in VanPatten / Williams (2015a). Einen historischen Abriss von Ansätzen, die die Verarbeitung einer Zweitsprache im Fokus haben und den angelsächsischen Forschungskontext einbeziehen, liefern Pienemann / Johnston (1996). Zu jüngeren Studien im Anschluss an diese ‚klassische' Phase von Pionierstudien in Deutschland vgl. das Kapitel 2.4 zum Forschungsstand.

4 Einerseits ist die Bedeutung dieser frühen Studien für die Theoriebildung und die didaktische Diskussion im Fachkontext DaF/DaZ kaum zu überschätzen, da durch sie zum ersten Mal umfangreiche und sehr detaillierte „Sprachprofile für einzelne Lerner bestimmter L1-L2-Konstellationen gewonnen" (Ahrenholz 2020b: 108) werden konnten. Die Modellierung dieser unterschiedlichen Sprachprofile hatte eine enorme Wirkung auf spätere Ansätze und auf didaktische Grundlagen im Kontext DaF/DaZ (vgl. hierzu Kapitel 2.4.5). Andererseits sind diese Studien gemessen an derzeit gültigen Standards in der empirischen und genauer in der quantitativen Forschung „eher als mehr oder weniger umfassende Fallstudien zu betrachten, deren Verallgemeinerbarkeit problematisch erscheint" (Ahrenholz 2020b: 108).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Für das Deutsche als Zweitsprache hat sich in diesen und in darauffolgenden Ansätzen zur Untersuchung syntaktischer Strukturen von Lernenden die Wortstellung als relevanter Untersuchungsgegenstand etabliert und viele der ‚klassischen' Studien analysieren die Wortstellung in Lernersprachen als wichtigen Indikator für die zweitsprachliche Entwicklung (vgl. Ahrenholz 2012a; Gaillard / Tremblay 2016: 420). Die sich entwickelnde Syntax einer Lernersprache ist, was die produktive Seite angeht, relativ gut nachvollziehbar und beschreibbar. Dies hängt mit der für das Deutsche spezifischen Klammerstruktur zusammen (vgl. Dürscheid 2012: 90), die mithilfe des topologischen Satzmodells (vgl. Wöllstein 2010: 20) bzw. dem Stellungsfeldermodell (vgl. Dürscheid 2012: 88) präzise dargestellt werden kann.[5] Das topologische Satzmodell ermöglicht eine genaue Beschreibung der unterschiedlichen Wortstellungsmuster (bzw. Satztypen) des Deutschen. Ein wesentliches Merkmal zur Beschreibung unterschiedlicher Wortstellungsmuster ist die Stellung des Verbs. Verben können je nach realisiertem Muster in Erststellung (Beleg 1), in Zweitstellung (Beleg 2) oder in Endstellung (Beleg 3) auftreten (vgl. Tabelle 1). Die Bezeichnung der Stellung bezieht sich auf das jeweilige topologische Feld. Die angeführten Belege verdeutlichen einige grundsätzliche Variationsmöglichkeiten bei der Besetzung. Das Vorfeld kann beispielsweise wie in Beleg 1 unbesetzt bleiben oder durch das Subjekt des Satzes besetzt werden wie in Beispiel 2. Darüber hinaus kann das Vorfeld mit einem anderen Element als dem Subjekt besetzt werden, was die Inversion von Prädikat und Subjekt zur Folge hat und dafür sorgt, dass das Subjekt im Mittelfeld auftritt. Dies ist in Beleg 4 der Fall. Einteilige Prädikate wie in Beleg 2 besetzen die linke Satzklammer. Zweiteilige Prädikate und Perfektformen besetzen die linke und die rechte Satzklammer. Nebensätze mit Verbendstellung können in komplexen Sätzen entweder im Vorfeld oder im Nachfeld auftreten. Werden sie, wie in Beleg 3, isoliert betrachtet, besetzt die nebensatzeinleitende Konjunktion die linke Satzklammer, während das finite Verb die rechte Satzklammer bildet.

	Vorfeld	Linke Satzklammer	Mittelfeld	Rechte Satzklammer	Nachfeld
1		Kann	ich gar nicht genau	sagen.	
2	Ich	weiß	genau,		dass das alles viel Zeit braucht.
3		dass	das alles viel Zeit	braucht.	
4	Jetzt	bin	ich sehr glücklich,		weil es geschafft ist.

Tabelle 1: Schematische Übersicht des topologischen Satzmodells

Im Bereich der syntaktischen Zweitsprachentwicklung wurden für das Deutsche als L2 in zahlreichen Studien (vgl. dazu Kapitel 2.4.2) universelle und implikationelle Erwerbssequenzen herausgearbeitet, die sich bei allen Lernenden unabhängig von ihrer jeweiligen Erstsprache und trotz individueller Variation (Alter, Auswahl, 2010) strukturell ähnlich entwickeln (vgl. Pienemann 1998).

5 Ausführliche Darstellungen des topologischen Satzmodells bzw. der Felder im Deutschen finden sich bei Dürscheid (2012: 87–105) und in Gallmann (2016: 871–904).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Vorbildung etc.) hinsichtlich der Wortstellung einstellen. Pienemann bündelt die Ergebnisse einer Reihe von Studien in der Formulierung der *Processability Theory* (PT) (vgl. Pienemann 1998).[6] Aus einer anderen Perspektive werden im Gemeinsamen europäischen Referenzrahmen für Sprachen syntaktische Strukturen unter den Skalen zur grammatischen Kompetenz gefasst (vgl. Europarat 2001: 113–115. ‚Grammatische Kompetenz' wird dort als „Kenntnis der grammatischen Mittel einer Sprache und die Fähigkeit, diese zu verwenden" (ebd.: 113) beschrieben. Kompetenz ist sicherlich stark kontextabhängig und zeitlich nicht stabil. Die Einschätzung darüber, was in welchem Kontext korrekt bzw. angemessen ist, könnte von verschiedenen Beurteilenden zu verschiedenen Zeitpunkten sehr unterschiedlich ausfallen. Grammatische und spezifischer syntaktische Strukturen, die heute als unmarkiert gelten, könnten vor einigen Jahren als stark markiert aufgefallen sein (vgl. van Dijk / Verspoor / Lowie 2011: 56). Kompetenz ist daher in diesem Kontext eher als variable Bestimmung der Fähigkeit von Lernenden zu verstehen, bestimmte Aufgaben in spezifischen Kontexten angemessen erfüllen zu können. Zur Syntax findet sich im GER folgende Anmerkung:

> Die ***Syntax*** behandelt die Anordnung von Wörtern zu Sätzen gemäß den beteiligten Kategorien, Elementen, Klassen, Strukturen, Prozessen und Beziehungen, die oft als Regelmenge repräsentiert werden. Die Syntax der Sprache eines erwachsenen Muttersprachlers ist sehr komplex und weitgehend unbewusst. Die Fähigkeit, Sätze zu bilden, um einen Sinn zu vermitteln, stellt einen zentralen Aspekt der kommunikativen Kompetenz dar (ebd.: 115; [Kursivsetzung im Original MS]).

Neben der lernersprachlichen Syntaxentwicklung existieren weitere Bereiche, die für die Erforschung von Lernersprache relevant sind. Zu nennen sind hier die Teilbereiche Lexik und Wortschatzentwicklung, Morphologie, Pragmatik und Phonetik/Phonologie (vgl. Petersen 2014: 8; Ahrenholz 2014: 169). Insgesamt kann davon ausgegangen werden, dass all diese Teilbereiche in komplexen und dynamischen Beziehungen zueinanderstehen und miteinander verbunden sind und dass grundsätzlich fraglich ist, inwiefern die Untersuchung eines dieser Teilbereiche Erkenntnisse über die zweitsprachliche Entwicklung als Ganzes liefern kann (vgl. Larsen-Freeman 2007; Bot / Lowie / Verspoor 2007). Diese Teilbereiche könnten zunächst als Teilsysteme analysiert werden, die wiederum konzeptionell in einen größeren Entwicklungskontext eingebettet werden können. Ziel hierbei könnte sein, grundlegende Parameter zu identifizieren, die die dynamische Verbundenheit verschiedener zweitsprachlicher Entwicklungsprozesse sichtbar machen (Larsen-Freeman 2007: 37).

6 Bereits vor der Konzeptualisierung der PT wurde die Frage aufgeworfen, inwiefern auf Erwerbssequenzen durch Unterricht bzw. durch gezielten Input Einfluss genommen werden kann (zur *teachability theory* Pienemann 1984). Ein kurzer Überblick zu Studien, die sich mit den Implikationen der *teachability theory* beschäftigen, findet sich in Grießhaber (2001a).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

In den letzten Jahren gewann im Kontext der Erforschung von Lernersprache außerdem die Korpuslinguistik als linguistischer Teilbereich im Forschungskontext DaF/DaZ, trotz teilweiser begrifflicher Uneinigkeit darüber, was letztlich als Korpus zählt, so sehr an Bedeutung (vgl. Meißner / Lange / Fandrych 2016), dass sie heute als linguistische „Standardmethode" (Hirschmann 2019: 2) gelten kann.[7] Prinzipiell kann davon ausgegangen werden, dass es sich bei einem Korpus um „eine Sammlung von Textdaten, also Sprache im Kontext, [handelt], die dem Zweck der linguistischen Auswertung dient und eine quantitative Auswertung von (qualitativen) sprachlichen Merkmalen zulässt" (Hirschmann 2019: 2). Für Linguistinnen und Linguisten ist der Rückgriff auf Korpora deshalb so ergiebig, da die Arbeit mit Korpora den Einbezug und die Analyse einer großen Anzahl an vergleichbaren und sonst nicht ohne großen Aufwand zugänglichen Daten erlaubt. Dies gilt in besonderem Maße für den Bereich der Lernersprachenforschung. Selbst wenn Steinhoff (2017: 359) noch Defizite sieht, was die Verfügbarkeit und die Auswertung von lernersprachlichen Korpora angeht, wurden in den letzten Jahren umfangreiche und aufwendig annotierte lernersprachliche Korpora erstellt, von denen hier beispielhaft einige kurz genannt werden sollen.[8] Zu diesen großen Projekten zählen die Falko-Lernerkorpora (vgl. u.a. Lüdeling et al. 2008 und Reznicek et al. 2012), die sowohl Korpora mit Daten beginnender Lernender (etwa das KanDel-Korpus; Vyatkina 2016) als auch Daten fortgeschrittener DaF-Lernender (z.B. das Essay-Kernkorpus, das WHiG-Korpus oder das Zusammenfassungskorpus) beinhalten.[9] Außerdem ist das trilinguale MERLIN-Lernerkorpus zu nennen (vgl. Abel et al. 2014 und MERLIN-Projekt 2014).[10] Ein sehr aktuelles Beispiel bietet darüber hinaus das DISKO-Korpus (vgl. Wisniewski et al. 2022 und Muntschick et al. 2020) aus dem SpraStu-Projekt, in dem schriftliche Sprachdaten fortgeschrittener Lernender longitudinal erhoben wurden.[11] Im Bereich der Korpuslinguistik hat sich darü-

7 Für ein frühes Plädoyer zum systematischen Einbezug korpuslinguistisch-basierter Ansätze im Kontext Deutsch als Fremdsprache vgl. Fandrych / Tschirner (2007). Einschätzungen zur gewachsenen Bedeutung der Korpuslinguistik finden sich etwa bei Eichinger (2018) für Korpora im Allgemeinen und bei McEnery et al. (2019) für die Erforschung von Lernersprache mithilfe von Lernerkorpora im Speziellen. Eine umfassende Einführung in das Feld bietet Hirschmann (2019).

8 Hier sei darauf hingewiesen, dass Steinhoff (2017) alle Korpora als Lernerkorpora bezeichnet, in denen Daten von lernenden Personen gesammelt werden. Damit legt er einen weiteren Begriff an als etwa die Korpuslinguistik, die den Begriff vor allem auf L2-Korpora bezieht (vgl. ebd. 354).

9 Der Zugang und weitere Informationen zu den Falko-Korpora finden sich unter: https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko/design [23.02.2023].

10 Der Zugang und weitere Informationen zum MERLIN-Korpus finden sich unter: https://merlin-platform.eu/index.php [23.02.2023].

11 Der Zugang und weitere Informationen zum DISKO-Korpus des SpraStu-Projektes finden sich unter: https://home.uni-leipzig.de/sprastu/korpora/disko/

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

ber hinaus mittlerweile die Anwendung zahlreicher statistischer Verfahren etabliert (vgl. zum Beispiel Brezina 2018), die durchaus vor dem Hintergrund der grundlegenden Kritik an der Replizierbarkeit empirischer Studien in verschiedenen Fachbereichen (vgl. Ioannidis 2005) gesehen werden muss. Hier haben in den letzten Jahren Fragen der methodischen Qualität und der statistischen Verlässlichkeit vor allem im Kontext des *methodological turns* an Bedeutung gewonnen (vgl. u.a. Byrnes 2013; vgl. Plonsky 2014; vgl. Marsden / Plonsky; für Hinweise zu einer statistischen Neuorientierung vgl. Cumming 2014).[12] Gerade in der englischsprachigen SLA-Forschung sind diese Entwicklungen stark durch die Reformbewegung des *methodological turns* beeinflusst. Kernforderungen, die sich aus dieser methodischen Erneuerungsbewegung ergeben, sind unter anderem eine besser begründete Auswahl statistischer Verfahren, ein transparenteres und standardisiertes Vorgehen und eine größere Einheitlichkeit beim Bericht statistischer Ergebnisse (vgl. Larson-Hall / Plonsky, 2015). Für den deutschsprachigen Kontext DaF/DaZ steht die angestoßene, methodische Erneuerungsbewegung zumindest in Teilen noch aus.

1.2 Erkenntnisinteresse und Forschungsfragen

In der vorliegenden Studie werden die Produkte von Schreibprozessen in den Fokus genommen und hinsichtlich produzierter syntaktischer Strukturen untersucht. Schreiben und die dadurch entstehenden Produkte sind in der Regel in institutionelle Bildungskontexte eingebunden und werden in diesen gefördert und gefordert (vgl. Ortega 2015: 85; vgl. Byrnes 2020: 245).[13] Die Analyse von zweitsprachlicher Entwicklung und dem Unterricht, in den diese eingebunden ist, sollten als ineinander verschränkt gesehen werden. Ortega (2015) spricht daher von ‚instruierter Entwicklung' (ebd.: 85; *instructed development*). Entwicklungsprofile von Lernenden in gesteuerten Lernkontexten können nicht ohne Rückgriff auf die speziellen Lehr- und Lernsituationen bzw. auf den curricularen Gesamtkontext analysiert und interpretiert werden (vgl. Vyatkina et al. 2015 30). Für Lernende bzw. für Schreibende ergibt sich in unterrichtlichen Kontexten die Notwendigkeit, sich in Reaktion auf angebotene bzw. von ihnen verlangte Schreibaufgaben linguistisch weiterzuentwickeln, um diese angemessen bearbeiten zu können. Bei dieser Weiterentwicklung vergrößern die Lernenden

[23.02.2023].

12 Während diese Kritik zunächst im Kontext medizinischer Studien geäußert wurde, hat sich die Überzeugung einer notwendigen Veränderung der Anforderungen an empirische Forschung mittlerweile auf die meisten wissenschaftlichen Fachgebiete ausgeweitet. Dies gilt ebenso für Studien, die sich im Bereich der Zweitspracherwerbsforschung verorten.

13 In diesem Zusammenhang genereller kann davon ausgegangen werden, dass selbst der mündliche Sprachgebrauch im Unterricht letztlich im Kontext „einer textgeprägten Sprache" (Schmölzer-Eibinger 2010: 1133) operationalisiert wird.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

ihr grammatisches Repertoire. Eine solche Erweiterung der Schreibfähigkeiten in einer L2 sollte sich in und an den produzierten Texten empirisch belegen lassen (vgl. Ortega 2015: 93).[14] Die Forschungsfragen zielen in der Folge darauf ab, herauszuarbeiten, wie und wann sich ausgewählte syntaktische Strukturen bei Lernenden entwickeln und ab wann und wie diese nachweisbar sind bzw. als erworben gelten können. Im spezifischen Kontext dieser Studie ergibt sich zusätzlich eine Forschungsfrage, die auf mögliche Effekte der experimentellen Manipulation innerhalb der Studie abzielt. Die vorliegende Studie versucht folgende drei Fragestellungen zu beantworten:

Erwerbsreihenfolgen:

- Welche syntaktische(n) Erwerbsreihenfolge(n) ergibt (ergeben) sich für erwachsene Deutschlernende mit Arabisch als L1? (Forschungsfrage 1)
- Wie lässt sich die syntaktische Entwicklung der gesamten Gruppe und von ausgewählten Teilnehmenden charakterisieren? (Forschungsfrage 2)

Instruktion:

- Lassen sich aufgrund der Instruktion bzw. der experimentellen Manipulation innerhalb der Studie (explizite Sprachvermittlung mit Schwerpunkt auf Syntax vs. implizite Vermittlung mit Schwerpunkt auf Semantik) Unterschiede zwischen den Teilnehmerinnen und Teilnehmern feststellen? (Forschungsfrage 3)

Methodologisch wird in der Arbeit durch eine Kombination bzw. eine Gegenüberstellung verschiedener methodischer Verfahren stets aufs Neue versucht, aufzuzeigen, welche Potenziale und welche Beschränkungen die ausgewählten methodischen Verfahren aufweisen können und warum es sinnvoll erscheint, Sprachdaten von Lernenden multimethodisch in den Blick zu nehmen. Es werden im Folgenden sowohl Gruppendaten als auch die Daten einzelner Lernender mithilfe von Einzelfallanalysen analysiert. Dabei werden methodische Verfahren wie longitudinale Clusteranalysen, visuelle Trendanalysen, verschie-

14 Hier sollte der analytische Schwerpunkt in einem anschließenden Schritt dann nicht nur auf den einzelnen Strukturen liegen, sondern darauf, wie diese Strukturen Texte herstellen und aufbauen und damit eine Perspektive „auf den sprachlichen Text" (Feilke 2015: 60) eröffnen. Die einzelnen Texte dienen dann als Indikatoren für die sich entwickelnde Schreibkompetenz eines Individuums (vgl. Steinhoff 2017: 353), wobei „Schreibkompetenz [gleichsam; MS] als Textkompetenz" (ebd: 353) aufgefasst werden kann. Der aus der Textlinguistik und Schreibforschung stammende Begriff der Textkompetenz (vgl. Schmölzer-Eibinger 2010: 1131) bezieht sich sowohl auf die Fertigkeit Texte zu lesen, zu kontextualisieren und kritisch mit dem eigenen Weltwissen in Verbindung zu setzen als auch darauf, Texte in angemessener bzw. adressatengerechter Weise zu verfassen (vgl. ebd.). Schreiben ist eine soziale und eine kommunikative Tätigkeit (vgl. Ortega 2015: 85; Steinhoff 2017: 353).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

dene Korrelationsanalysen und *Change Point*-Analysen herangezogen, die unter anderem darauf abzielen, innerhalb der Entwicklungsverläufe der Lernenden inter- und intraindividuelle Variabilität sichtbar zu machen. Gerade bei den longitudinalen Clusteranalysen und den *Change Point*-Analysen handelt es sich um Verfahren, die bisher nur in wenigen SLA-Studien und dort vor allem im Kontext der Modellierung von Motivation und m. W. in keinen Studien Anwendung finden, die sich auf das Deutsche als L2 beziehen.[15] In allen Analysen steht die Entwicklung der Lernenden und nicht etwa die Analyse von Fehlern oder Abweichungen innerhalb der Lernersprache, im Vordergrund (vgl. Kapitel 3.2 und 3.3). Ein weiteres Ziel dieser Arbeit ist, für Gruppen von Lernenden und für einzelne Lernende mithilfe sehr dichter Datensets anhand ausgewählter Variablen herauszuarbeiten, wie sich verschiedene syntaktische Variablen im Laufe der Zeit gegenseitig beeinflussen und dabei die (nichtlineare) Natur von zweitsprachlichen Entwicklungsprozessen ausmachen.

Eine Grundeinheit zur Beschreibung der schriftlichen Daten bildet in der vorliegenden Untersuchung die „minimale satzwertige Einheit (MSE)" (Grießhaber 2005). Auf der Basis von minimalen satzwertigen Einheiten werden jeweils weitere Variablen, wie etwa die durchschnittliche Länge einer MSE berechnet. Außerdem wurden für alle Texte weitere Variablen berechnet, die sich auf die Subordination oder das Vorkommen komplexer Nominalphrasen beziehen (vgl. zur Auswahl der Variablen und Messeinheiten Kapitel 3.1). Gleichzeitig sollen die ausgewählten Variablen so modelliert werden, dass sie es ermöglichen, die Prozesse und dynamischen Interaktionen sichtbar zu machen, die im Verlauf der (schriftlichen) Zweitsprachentwicklung zwischen diesen Variablen existieren.

1.3 Hybride Lern- und Erwerbssituation zwischen DaF und DaZ

Die Aneignung und das Lernen von Sprachen sind grundsätzlich in einem Kontext lebensweltlicher Mehrsprachigkeit (Gogolin 2004) zu verorten und können nicht isoliert von diesem Kontext gesehen werden. Es kann vielmehr von einer komplexen und dynamischen „Verschränkung der Sprachen miteinander ausgegangen werden" (Knappik / Dirim 2018: 170). Eine solche Konzeptualisierung von Sprachen und Sprachenlernen lässt die „Einteilung der Sprachen in Erstsprache(n) und Zweitsprache eher künstlich" (Knappik / Dirim 2018: 170) erscheinen. In ähnlicher Weise wird in jüngster Zeit die Dichotomie zwischen

15 Vgl. zur Anwendung longitudinaler Clusteranalysen in Bezug auf die Motivation von Studierenden im Laufe eines Semesters Pinel / Csizér (2015). Baba / Nitta (2014) wenden in ihrer Untersuchung zu sich verändernden Selbsteinschätzungen während des Erwerbs einer L2 *Change Point*-Analysen an; vgl. zu beiden Studien die Kapitel 2.5.2 zum Forschungsstand und Kapitel 3.2.5 und 3.3.5 zum methodischen Vorgehen).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Deutsch als Fremdsprache und Deutsch als Zweitsprache verstärkt diskutiert und zunehmend in Frage gestellt (vgl. etwa Zabel et al. 2022; vgl. Riehl / Schroeder 2022).[16]

Für eine Verortung der vorliegenden Studie im Kontext von Deutsch als Fremd- oder Zweitsprache (DaF/DaZ) sind neben diesen konzeptionellen Überlegungen einige Punkte zu beachten (vgl. für die folgenden Ausführungen Ahrenholz 2020a: 13). Eine grundsätzliche definitorische Überlegung bezieht sich hier auf das Begriffspaar ‚Lernen und Erwerben', wobei ‚Lernen' oft mit strukturiertem Input innerhalb eines Sprachkurses, also vermittelt durch Lehrkräfte und auf der Grundlage von Lehrwerken, verbunden wird (vgl. z. B. Hufeisen 2003: 97; Ahrenholz 2020a). ‚Erwerben' bezeichnet im Gegensatz dazu einen eher ungesteuerten Prozess, der hauptsächlich durch die Aneignung von sprachlichen Strukturen in tatsächlichen Kommunikationssituationen und ohne direkten Rückgriff auf die Vermittlung durch eine Lehrkraft oder ein Lehrwerk gekennzeichnet ist (vgl. Hufeisen 2003: 97; vgl. Hünlich et al. 2018: 12). Obwohl die Teilnehmenden der vorliegenden Studie in einem immersiven Umfeld Deutsch lernen, welches als zweitsprachliches Umfeld gekennzeichnet ist, nämlich in Leipzig in einem deutschsprachigen Kontext, kann aufgrund des intensiven Charakters des Kurses davon ausgegangen werden, dass wesentliche sprachliche Aneignungsprozesse im Unterricht stattgefunden haben. Gleichzeitig ist der Erwerb der deutschen Sprache durch die Fluchtgeschichten der Teilnehmenden geprägt. Die Lernerinnen und Lerner standen vor der Herausforderung, in Deutschland „hochrelevante kommunikative Aufgaben mit eventuell unzureichenden Sprachkompetenzen" (Hünlich et al. 2018: 12) bewältigen zu müssen. Der Beginn des Spracherwerbs fällt bei den Teilnehmenden mit dem Beginn des Aufenthalts in Deutschland zusammen. Zuletzt war der Unterricht in den Intensivkursen so konzipiert, dass Rückgriffe auf die L1 Arabisch vor allem in der Anfangsphase des Kurses möglichst vermieden werden sollten. Dazu kam, dass nicht alle Lehrkräfte Kenntnisse im Arabischen besaßen. Trotzdem war Arabisch zusätzlich dadurch als Umgangssprache präsent, da alle Teilnehmenden Arabisch als Erstsprache sprachen und darüber hinaus dauerhaft arabischsprachige Studienassistentinnen und Studienassistenten den Ablauf in der Studie unterstützten. Die Teilnehmenden waren folglich nicht notwendigerweise darauf angewiesen, bei Fragen zu administrativen bzw. allgemeinen Abläufen der Studie auf Deutsch zurückzugreifen. Gleichzeitig ist bei der Gruppe von Teilnehmenden mit einem sehr hohen Interaktionsgrad mit deutschen Behörden zu rechnen, da es sich bei allen Teilnehmenden um Personen handelte, die entweder einen Flüchtlingsstatus gemäß der Genfer Flüchtlingskonvention besaßen oder einen subsidiären Schutz zugesprochen bekommen hatten, weil keine politi-

16 Für eine weitere Problematisierung der starken Dichotomie des Begriffspaars DaF/DaZ und den Zuschreibungen gesteuert/ungesteuert vgl. Rösler (2020). Rösler gibt in seinem Beitrag sehr einschlägige Beispiele für Fälle, in denen dichotome Zuordnungen wenig Sinn ergeben oder kontraproduktiv sein könnten.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

sche Verfolgung im Sinne des Art. 16a GG vorlag und die Voraussetzungen des Flüchtlingsschutzes gem. § 3 AsylVfG nicht gegeben waren (vgl. Kapitel 3.1.2).

Es kann festgehalten werden, dass die sprachlichen Entwicklungsprozesse der Teilnehmenden innerhalb der Studie als weitgehend zweitsprachlich gekennzeichnet werden können, dass aber durchaus Bereiche existieren, in denen die Entwicklungsprozesse und die Lehr-Lernkontexte eher fremdsprachliche Charakteristika aufweisen.

1.4 Abgrenzung der vorliegenden Arbeit zur Gesamtstudie

Das Wissen um neuronale Grundlagen des menschlichen Gehirns ist in den letzten Jahrzehnten kontinuierlich erweitert worden (Friederici 2011: 1357; Berwick / Friederici / Chomsky / Bolhuis 2013:89). Erst seit wenigen Jahren ermöglichen außerdem bildgebende Verfahren (etwa die Magnetresonanztomographie (MRT) und die funktionelle Magnetresonanztomographie (fMRT)), erhobene Spracherwerbsdaten systematisch mit neurowissenschaftlichen Aufnahmen des menschlichen Gehirns zu korrelieren.[17] In den letzten Jahren wurde außerdem in diversen neurowissenschaftlichen Studien mit Erwachsenen der Einfluss von (Zweit- und Fremd-)Spracherwerb auf die neuronalen Netzwerke des Gehirns gezeigt (vgl. hierzu und für einen Überblick verschiedener Studien Li et al. 2014; Abutalebi / Green 2007; Hernandez 2013).[18] Goucha et al. (2015) konnten in einer

17 Die relevanten und bis heute stark rezipierten Spracherwerbsstudien im deutschsprachigen Raum fallen zu einem großen Teil in eine Zeit, in der sie noch nicht durch genaue bildgebende Verfahren gestützt werden konnten. Außerdem fällt auf, dass Veröffentlichungen im Bereich DaF/DaZ kaum Erkenntnisse der Neurowissenschaft systematisch einbeziehen und neurowissenschaftliche Veröffentlichungen nur selten explizit auf Ansätze der Spracherwerbsforschung der letzten Jahre eingehen. Viele der neurowissenschaftlichen Studien zum Zweitspracherwerb testen Zweitspracherwerb im Rahmen sehr enger experimenteller Parameter und mit Testformaten, die der Komplexität und der Dynamik des tatsächlichen Zweitspracherwerbs kaum gerecht werden können (vgl. Selinker 1972; Selinker 2014).

18 Neben dem Einfluss des Lernens von Sprachen auf die neuronalen Netzwerke wurde vor allem in der Psycholinguistik seit längerer Zeit der Einfluss individueller, neuronaler Voraussetzungen auf das potenzielle Sprachlernen untersucht. Die linguistischen Kompetenzen eines Lernenden scheinen mit seiner Verarbeitungseffizienz (van Moere 2012) zusammenzuhängen. Zentral für die effektive Verarbeitung von sprachlichem Input sind die Konzepte von Automatisierung und Arbeitsgedächtnis. Mit Automatisierung ist vor allem die Leichtigkeit gemeint, mit der gesprochene Sprache verstanden und produziert wird (Segalowitz 2003). Leichtigkeit steht hier für den Grad der Prozeduralisierung von Regeln, d.h., den Übergang von deklarativem zu prozeduralem Wissen. Das Arbeitsgedächtnis bezeichnet das Konzept der zeitlich begrenzten Speicherung und Bearbeitung von Informationen, die für eine Reihe kognitiver Hand-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

sprachvergleichenden Studie zeigen, dass sich die Struktur des Sprachnetzwerkes bei Sprecherinnen und Sprechern unterschiedlicher L1 signifikant unterscheidet.[19] Die Ergebnisse dieser Studien wurden im Kontext der deutschsprachigen Zweitspracherwerbforschung kaum wahrgenommen, was damit zusammenhängen könnte, dass es sich bei neurowissenschaftlichen Studien in der Regel um Experimente und damit nicht um ‚natürliche' Zweitspracherwerbsprozesse handelt. Seit einiger Zeit gewinnt das Konzept der Plastizität des Gehirns in den Neurowissenschaften an Bedeutung. Es konnte nachgewiesen werden, dass sich adulte neuronale Netzwerke noch an neue Stimuli und Lernkontexte anpassen können und Veränderungen im Gehirn nicht mit dem Abschluss der Pubertät ausbleiben (siehe hierzu etwa Draganski et al. 2004; Taubert et al. 2010 und Taubert et al. 2012). Dies sind Ergebnisse, die gerade im Kontext der Fremd- und Zweitspracherwerbsforschung von hoher Relevanz sind, da sie die Diskussion um kritische Perioden in Erwerbsprozessen neu perspektivieren könnten.[20]

Die Gesamtstudie am Max-Planck-Institut für Kognitions- und Neurowissenschaften (MPI CBS) und am Herder-Institut (HI), aus der die Daten der vorliegenden Studie stammen, ist konzeptuell in diesen größeren neurowissenschaftlichen Forschungskontext eingebunden.[21] Einerseits bot die im Kapitel 1.1 zur Problemstellung und Relevanz geschilderte Situation aus neurowissenschaftlicher Sicht die Chance, in großem Umfang longitudinale anatomische und funktionelle Gehirndaten arabischer L1-Sprechender zu erheben, die in dieser

lungen relevant sind (Baddeley 2003: 1989). Das Arbeitsgedächtnis wird daher einerseits als relevant für Sprachproduktion und Sprachverarbeitung gesehen, andererseits jedoch als sprachunabhängig (Gaillard / Tremblay 2016:421). Aus neurowissenschaftlicher Sicht ist darüber hinaus das Konzept relevant, dass für das Lernen einer Sprache das deklarative und das prozedurale Gedächtnis von zentraler Bedeutung sind (vgl. Ullman 2016; Ullman / Lovelett 2018). Deklaratives Gedächtnis wird und wurde immer wieder in Zusammenhang mit dem Lernen expliziter Inhalte gebracht, während das prozedurale Gedächtnis häufig mit dem Verarbeiten impliziten Wissens oder von Fähigkeiten, die bereits automatisiert wurden, in Verbindung gebracht wird (vgl. Ullman 2016: 3).

19 Eine wesentliche Grundlage für die Definition des Sprachnetzwerkes als Netzwerk ist der Informationsaustausch verschiedener Regionen des Gehirns, die in die Verarbeitung von Sprache involviert sind. Für einen Überblick über die verschiedenen Komponenten des Sprachnetzwerkes siehe Friederici (2011) und Friederici (2013).

20 Für eine ausführliche und differenzierte Zusammenfassung der Diskussion um kritische Perioden im Spracherwerb aus Sicht der Neurowissenschaften siehe Friederici (2017: 145–162). Für eine ausführliche Diskussion aus der Perspektive des ungesteuerten Zweitspracherwerbs vgl. Czinglar (2014: 5–29). Eine kritische Perspektive aus methodischer Sicht regen Lowie / Verspoor (2015: 79) an.

21 Für einen Überblick über das Projekt vgl. die Projektseite auf der Homepage des Max-Planck-Instituts für Kognitions- und Neurowissenschaften: https://www.cbs.mpg.de/neuropsychologie/zweitspracherwerb [23.02.2023].

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Form bis jetzt nicht verfügbar waren.[22] Andererseits wurden „Veränderungen in der strukturellen Verbindungsstärke zwischen den Sprachregionen [im Laufe des Zweitspracherwerbs einer Sprache; MS] bisher noch nicht untersucht" (Anwander / Friederici 2015: 2). Die Fragestellungen und Erkenntnisinteressen der Gesamtstudie unterscheiden sich in der Folge zum Teil relativ deutlich von den Erkenntnisinteressen der vorliegenden Arbeit.[23]

Während in der Gesamtstudie die Plastizität des Gehirns beim Erwerb der Zweitsprache im Fokus steht, geht die vorliegende Studie explizit der Frage nach, wie sich die Entwicklung von syntaktischen Kompetenzen beobachten, beschreiben und letztlich modellieren lässt. Dabei werden für die vorliegende Studie nur Sprachdaten und (behaviorale) Hintergrunddaten verwendet, da der Verfasser für die Erhebung und Auswertung dieser Daten direkt verantwortlich war. Das bedeutet gleichzeitig, dass in dieser Arbeit weder neurowissenschaftliche Daten ausgewertet noch die herausgearbeiteten Ergebnisse direkt auf diese bezogen werden. Die Ergebnisse dieser Studie wieder auf die neurowissenschaftlichen Ergebnisse der Gesamtstudie zu beziehen, bleibt daher ein Desiderat weiterer Forschungsschritte.

1.5 Aufbau der Arbeit

Die vorliegende Arbeit ist im Anschluss an diese Einleitung in vier weitere Kapitel gegliedert, die im Folgenden kurz vorgestellt werden.

In Kapitel 2 ‚Forschungsstand' werden die Grundlagen verschiedener Ansätze der Zweitspracherwerbsforschung vorgestellt, die zur Beantwortung der Forschungsfragen dieser Untersuchung herangezogen werden. Dabei wird zunächst ein Blick auf die Unterschiede und Gemeinsamkeiten von Erst- und Zweitspracherwerb geworfen (Kapitel 2.1.) und dann beschrieben, welche Ansätze sich zur Erforschung und Analyse von Lernersprache herausgebildet haben (Kapitel 2.2). Darauf aufbauend werden in Kapitel 2.3 Ansätze der deutschsprachigen Schreibforschung vorgestellt. Die beiden abschließenden Kapitel in diesem Teil der Arbeit widmen sich zwei unterschiedlichen Ansätzen der Zweitspracherwerbsforschung. Diese sind die *Processability Theory* (2.4) und die *Complex Dynamic Systems Theory* (2.5).

Kapitel 3 ‚Methodisches Vorgehen' ist den methodischen Grundlagen dieser

22 Dies hängt vor allem damit zusammen, dass in den arabischen Ländern in den letzten Jahrzehnten kaum Grundlagenforschung betrieben wurde bzw. betrieben werden konnte. Der Schwerpunkt der Forschung im arabischen Raum lag in den letzten Jahrzehnten fast ausschließlich im Bereich von Energie- und Wassertechnologie (vgl. Guessoum 2012).

23 Vgl. hierzu weitere Arbeiten aus der Gesamtstudie, unter anderem Wei et al. (2023).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Studie gewidmet und es werden die verschiedenen Analyseverfahren erklärt, die im Rahmen dieser Studie eingesetzt werden. Dabei werden in Kapitel 3.1 die Grundlagen der Studie hinsichtlich der gebildeten Stichprobe, des zeitlichen Ablaufs der Studie, der experimentellen Manipulation des Unterrichts und möglicher Limitationen besprochen. In der Studie werden Daten auf Gruppenebene und in Einzelfallanalysen mit zwei Geschwistern ausgewertet. Kapitel 3.2 beinhaltet das Vorgehen auf Gruppenebene und stellt die in diesem Zusammenhang gebildete Teilstichprobe sowie die eingesetzten methodischen Verfahren (wie die Emergenzanalysen und die longitudinalen Clusteranalysen) vor. In Kapitel 3.3 geschieht dies in ähnlicher Weise in Bezug auf die Einzelfallanalysen. Hier werden zunächst die Teilnehmenden und dann die Verfahren der Trendanalysen, der unterschiedlichen Korrelationsanalysen und der *Change Point*-Analysen vorgestellt.

In Kapitel 4 ‚Ergebnisse' werden darauffolgend die Ergebnisse der verschiedenen Analyseschritte vorgestellt und zusammengefasst, um in Kapitel 5 ‚Diskussion und Ausblick' im Kontext bisheriger Forschungserkenntnisse genauer diskutiert und kontextualisiert zu werden.

Im Anhang dieser Arbeit finden sich die eingesetzten Sprachtests, Metadaten zu allen Lernenden und die aus den Lernertexten extrahierten Strukturen und Komplexitätsmaße, die die Grundlage für die in dieser Studie durchgeführten Analysen bilden. Die Lernertexte selbst können derzeit auf Anfrage zur Verfügung gestellt werden. Zudem ist eine Veröffentlichung als eigenständiges und aufbereitetes Korpus geplant.

Der Anhang befindet sich frei zugänglich unter folgendem Link:
https://anhang-syntaktische-strukturen.ESV.info.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

2 Forschungsstand

In den folgenden fünf Kapiteln werden die für die vorliegende Arbeit ausschlaggebenden Theoriestränge und Perspektiven auf den Komplex des Zweitspracherwerbs bzw. der Zweitsprachentwicklung vorgestellt. Hierbei sollen wesentliche Begriffe und Konzepte erklärt und Verbindungslinien zu unterschiedlichen Anwendungskontexten gezogen werden. Da es sich bei der vorliegenden Studie um eine Zweitspracherwerbsstudie handelt, sollen in Kapitel 2.1 die wesentlichen Charakteristika des Zweitspracherwerbs auf der Grundlage von Unterschieden zum Erstspracherwerb vorgestellt werden. In Kapitel 2.2 wird eine sehr knappe Einführung in wesentliche Perspektiven bei der Beschreibung lernersprachlicher Äußerungen gegeben, wobei der Fokus auf der CAF-Beschreibung liegt (*complexity*, *accuracy*, *fluency*). Kapitel 2.3 lenkt den Fokus auf die Erforschung schriftlicher lernersprachlicher Produkte und legt einen Schwerpunkt auf den deutschsprachigen Forschungskontext DaF/DaZ. In den letzten beiden Unterkapiteln werden zwei Ansätze vorgestellt, die Zweitspracherwerb bzw. zweitsprachliche Entwicklung sehr unterschiedlich konzeptualisieren: die *Processability Theory* (Kapitel 2.4) und die *Complex Dynamic Systems Theory* (Kapitel 2.5). Ziel der folgenden Kapitel ist, jeweils eine kompakte Beschreibung der Grundannahmen und der theoretischen Hintergründe der jeweiligen Ansätze zu liefern, um diese miteinander in Beziehung setzen zu können. In diesem Kontext sollen Limitationen der jeweiligen Ansätze abgewogen werden. Darüber hinaus wird der Blick auf die wesentlichen Entwicklungen gelenkt werden, die diese Ansätze in den letzten Jahren durchlaufen haben.[24]

2.1 Unterschiede zwischen Erst- und Zweitspracherwerb

Im Folgenden sollen zunächst Unterschiede zwischen Erst- und Zweit- bzw. Fremdspracherwerb und solche erstsprachlichen Entwicklungsschritte im Fokus stehen, die sich auf die Entwicklung grammatischer bzw. syntaktischer Fertigkeiten beziehen. Hierbei kann nicht im Einzelnen auf sämtliche komplexe und vielfach miteinander verknüpfte Prozesse des Erstspracherwerbs eingegan-

24 Eine solche Fokussierung auf zwei ausgewählte theoretische Ansätze zur Beschreibung von fremd- bzw. zweitsprachlichen Entwicklungen stellt notwendigerweise eine starke Verkürzung dar und lässt viele weitere theoretische Perspektiven außer Acht. Ausführliche und aktuelle Einführungen und Überblicksdarstellungen zu Zweitspracherwerbstheorien bzw. zu SLA-Ansätzen und wesentlichen Hypothesen finden sich beispielsweise bei Hufeisen / Riemer (2010), Königs (2010), VanPatten / Williams (2015), Grießhaber (2016), Ahrenholz (2020b), und Bordag / Opitz (2021). Es kann davon ausgegangen werden, dass sich die zahlreichen unterschiedlichen Ansätze in der Auseinandersetzung miteinander entwickelt haben und dies weiterhin tun (vgl. Königs 2010: 755).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

gen werden, weswegen an geeigneter Stelle jeweils auf weiterführende und zusammenfassende Literatur verwiesen wird.

Während die definitorische Abgrenzung von Fremd- und Zweitspracherwerb in jüngster Zeit verstärkt in Frage gestellt wird (vgl. Kapitel 1; vgl. Zabel et al. 2022; vgl. Riehl / Schroeder 2022), lassen sich beide Erwerbs- bzw. Entwicklungstypen scheinbar relativ deutlich vom Erstspracherwerb abgrenzen und durch Unterschiede zu diesem definieren. Der Erstspracherwerb dient dem Zweit- und Fremdspracherwerb in vielerlei Hinsicht als Referenzpunkt (vgl. Meisel 2011: 13; für eine ausführliche Diskussion unterschiedlicher Erwerbskontexte vgl. Meisel 2021). Konsens besteht darüber, dass Erstspracherwerb in der Regel kindlicher Erstspracherwerb ist und damit an bestimmte, sehr frühe Entwicklungsphasen von Babys und Kleinkindern geknüpft und durch bestimmte kritische Perioden eingeschränkt wird (vgl. z. B. Meisel 2011: 202–239). Gleichzeitig scheint der Erstspracherwerb im Vergleich zu anderen Erwerbstypen relativ schnell und auf natürliche Weise abzulaufen, was bedeutet, dass er nicht durch explizites Lernen unterstützt werden muss (vgl. Meisel 2011: 240–241).[25] Dazu kommt, dass der Erstspracherwerb eines Kindes in Abhängigkeit des gegebenen Inputs für jede natürliche Sprache vergleichbar sowie in zum Teil invarianten Sequenzen abläuft und während der Schwangerschaft beginnt (vgl. Friederici 2017: 163).[26]

Grießhaber geht von grundsätzlichen Erwerbsschritten in den ersten 24 Monaten aus (vgl. Grießhaber 2016: 10). Tracy (2007) setzt in diesem Kontext zweieinhalb bis drei Jahre an (Tracy 2007:69) und Friederici (2017) identifiziert die ersten drei Lebensjahre als Phase, in denen Menschen besonders offen gegenüber dem Lernen einer Sprache sind und nach denen wesentliche Erwerbsprozesse in der zu erlernenden Erstsprache abgeschlossen sind (vgl. Friederici 2017: 183). Unklar bleibt nach wie vor, ob diese Phase eine kritische Periode oder eine sensitive Phase darstellt. Während der Begriff ‚kritische Periode' eine

25 Grießhaber (2016) geht davon aus, dass die Vermutung, Erstspracherwerb brauche keine externe Unterstützung, auf einer Reihe von terminologischen Verkürzungen beruhe. Vielmehr wird gerade Kindern vielfach lernförderlicher Input, wie Sprachspiele, zur Verfügung gestellt (vgl. Grießhaber 2016: 52–53). Viele der Vertreter/-innen universalgrammatischer Ansätze gehen davon aus, dass Kinder bei ihrer Sprachentwicklung Zugang zu einer angeborenen Universalgrammatik haben, die über eine *Language Acquisition Device* (LAD) gesteuert wird (vgl. Pienemann 1998; Meisel 2011; Chomsky 1965). Im Gegensatz dazu müssten erwachsene L2-Lernende beim Lernen einer weiteren Sprache verstärkt auf Verarbeitungsstrategien zurückgreifen (vgl. Pienemann 1998: 12).

26 Im Zusammenhang mit dem angebotenen Input wird oft vom ‚logischen Problem' des Erstspracherwerbs gesprochen. Dieser Ausdruck bezeichnet das Paradox, dass Kinder in einer relativ kurzen Zeit mit nur begrenztem Input viele der grundlegenden Prinzipien ihrer L1 erwerben, ohne dass alle diese Prinzipien aus den durch den Input gemachten Beobachtungen abgeleitet werden könnten (vgl. z. B. Pienemann 1998; Pienemann / Lenzing 2015: 161 und Kapitel 2.4).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Zeitspanne beschreibt, in der bestimmte Lernerfahrungen stattfinden müssen, um Lernen zu ermöglichen, zeichnet sich eine sensitive Phase dadurch aus, dass Lernerfahrungen in dieser Phase besonders förderlich sind und leichter aufgenommen werden können, jedoch nicht explizit auf diese Phase beschränkt sind (vgl. Czinglar 2014: 15). Friederici (2017) merkt etwa einschränkend an, dass die Belege für eine kritische Periode des Lernens aus neurowissenschaftlicher Perspektive bis heute unzureichend seien, wenn auch mittlerweile sehr deutlich sei, dass Lernen sich mit zunehmendem Alter verändere (vgl. Friederici 2017: 145).[27] Oft wird hier etwa der Beginn der Pubertät als relevante Schwelle angesetzt, die erstsprachlichen und frühen zweitsprachlichen Erwerb vom Zweitspracherwerb Erwachsener trennt (vgl. Czinglar 2014: 10).

Unabhängig von unterschiedlichen Geschwindigkeiten bei der Entwicklung bestimmter Strukturen, scheinen sich für fast alle Menschen musterhafte Erwerbssequenzen im erstsprachlichen Erwerb unterschiedlicher Strukturen zu zeigen (vgl. Friederici 2017: 163; siehe unten). Diese gehen mit der Beobachtung einher, dass sich der Erstspracherwerb durch eine geringere interindividuelle Variation als etwa der Zweit- oder Fremdspracherwerb auszeichnet. Eine geringere Variation in interindividuellen Entwicklungsverläufen kann dafür sorgen, dass sich letztlich ein homogeneres Bild ergibt, wenn es um sprachliche Kompetenzen von L1-Sprechenden geht (vgl. Czinglar 2014: 83). Selinker fasst zusammen, dass „Erstspracherwerb zu konvergenten Grammatiken führt, aber Zweitspracherwerb divergente Grammatiken zur Folge hat" (Selinker 2014: 225; [Übersetzung MS]).[28]

27 Friederici (2017) macht in ihrer Besprechung relevanter neurowissenschaftlicher Studien zum Zweitspracherwerb darauf aufmerksam, dass die Ergebnisse dieser Studien darauf hindeuten, dass die Fähigkeit zur Verarbeitung von Syntax eng mit dem Erwerbsalter zusammenhängt und bei späterem Erwerbsbeginn der weiteren Sprache (*age of acquisition* / *age of onset*) kontinuierlich abnimmt (vgl. Friederici 2017: 155). Meisel (2011) geht davon aus, dass Lernende, die eine Sprache nach ihrem dritten Lebensjahr lernen, die Syntax dieser Sprache nicht mehr auf einem erstsprachlichen Niveau erwerben bzw. beherrschen werden.

28 Es haben sich trotz großer Übereinstimmungen im Bereich der zeitlichen Entwicklungsdimension zahlreiche Erstspracherwerbstheorien herausgebildet, die erstsprachliche Entwicklung und die zugrundeliegenden Mechanismen unterschiedlich konzeptualisieren und erklären. Grießhaber (2016) etwa nennt den Behaviorismus, den Nativismus und den Interaktionismus, den Kognitivismus sowie den Konnexionismus als relevante Erklärungsansätze (vgl. Grießhaber 2016: 9–10). In dieser Arbeit kann leider nur sehr begrenzt auf die Theoriebildung in Bezug auf den Erstspracherwerb eingegangen werden. Ausführliche und zusammenfassende Erklärungen finden sich bei Grießhaber (2016). Aus einer universalgrammatischen Perspektive im Anschluss an Chomsky finden sich Erklärungen bei Meisel (2011: 2–61) und aus neurowissenschaftlicher Perspektive bei Friederici (2017: 145–200). Kompakte Hinweise zu unterschiedlichen Erwerbsphänomenen im Erst- und Zweitspracherwerb aus DaF/DaZ-Perspektive finden sich bei Byrant / Rinker (2021).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Es wird angenommen, dass die erstsprachliche syntaktische Entwicklung erst dann beobachtbar wird, wenn eine kritische Wortschatzmasse erreicht worden ist (vgl. Bryant / Rinker 2021: 276).[29] Analog dazu wird für das Lernen einer Zweitsprache, zumindest aus neurowissenschaftlicher Perspektive, zunächst eine anfängliche Phase angenommen, die mehr durch das Lernen von Wörtern geprägt ist und auf die eine Phase folgt, in der verstärkt grammatische Regeln als Lerngegenstand ins Zentrum rücken (vgl. Friederici 2017: 153; Opitz / Friederici 2004). Es wurden in diesem Zusammenhang unterschiedliche Modelle der erstsprachlichen, syntaktischen Entwicklung vorgeschlagen, die alle einen sequenziellen Verlauf der einzelnen Stufen annehmen. Clahsen (1982; 1986) schlägt für die syntaktische Entwicklung der Erstsprache Deutsch ein fünfstufiges Entwicklungssystem vor, das bei Kindern etwa im Alter von einem Jahr einsetzt. Tracy (2007a) und Tracy / Thoma (2009) arbeiten demgegenüber ein vierstufiges Verlaufsmodell für die Verbstellung heraus, das hier kurz angerissen werden soll.[30] Kinder im Alter von etwa einem bis anderthalb Jahren durchlaufen in diesem Modell zunächst eine Einwortphase, in der zum Teil einzelne Verbpartikeln realisiert werden (vgl. Tracy 2007a: 72). Auf diese Phase folgen erste Mehrwortäußerungen mit infiniten Verben am Ende der Äußerungen. Kinder befinden sich in der Regel im Alter zwischen anderthalb und zwei Jahren in dieser Phase (vgl. ebd.). An diese zweite Phase anschließend folgt eine Phase, in der zielsprachliche Sätze mit Verben in Zweitposition realisiert werden (vgl. ebd.). Diese Phase gilt erst als abgeschlossen, wenn zusätzlich Inversionen realisiert werden, bei denen das Subjekt nach dem Verb auftritt und das Vorfeld mit einem anderen Element als dem Subjekt besetzt wird. Sehr deutlich wird in dieser von Tracy (2007a) herausgearbeiteten Erwerbsprogression, dass „die Verbklammer von rechts nach links aufgespannt wird" (Bryant / Rinker 2021: 278). In einer vierten Phase, die Kinder im Alter zwischen drei und vier Jahren durchlaufen, werden Nebensätze mit Verben in Endstellung produziert (vgl. Tracy 2007a: 72). Dass der Erwerb der syntaktischen Wortstellungsregeln nicht immer innerhalb des aufgezeigten zeitlichen Fensters ablaufen muss und sich durchaus bis in das elfte Lebensjahr erstrecken kann und darüber hinaus durchaus heterogene und durch Variabilität gekennzeichnete Erwerbsverläufe bei erstsprachlichen Kindern zu beobachten sind, zeigt Ulrich (2017) in ihrer groß angelegten Querschnittsstudie mit 968 Kindern (vgl. Ulrich 2017: 67).[31]

29 Beobachtbar bezieht sich hier darauf, dass syntaktische Strukturen in den Äußerungen von Kindern zu identifizieren sind. Eine ausführliche Darstellung erstsprachlicher Wortschatzentwicklung findet sich bei Grießhaber (2016: 31–47).

30 Ein zusammenfassendes Modell beider Veröffentlichungen findet sich bei Czinglar (2014: 85).

31 Dass L1-Sprechende (bzw. -Lernende) durchaus nicht immer die homogene Gruppe bilden, als die sie bisweilen dargestellt oder konzeptualisiert werden bzw. dass die Annahme prototypischer L1-Sprechender unter Umständen insgesamt fallen gelassen werden könnte oder müsste, zeigen die Beiträge eines jüngst publizierten *Research Topic* der Zeitschrift *Frontiers in Psychology* „The Notion of

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Für die fremd- und zweitsprachliche Entwicklung lassen sich zahlreiche Faktoren herausstellen, die diese in anderer Weise beeinflussen als dies beim Erstspracherwerb der Fall ist. Diese Faktoren könnten folglich ausschlaggebend dafür sein, dass zweitsprachliche im Vergleich zu erstsprachlichen Entwicklungen durch ein größeres Maß an inter- und intraindividueller Variation geprägt sind. Zunächst ergibt sich aus der zeitlichen Abfolge von L1- und L2-Erwerb, dass sich die Anfangsbedingungen (*initial state*) der beiden Erwerbssituationen schon deshalb unterscheiden, weil im L2-Erwerb auf erstsprachliche und unter Umständen umfangreiche lebensweltliche Erfahrungen zurückgegriffen werden kann (vgl. Grießhaber 2018: 25–26). Gleichzeitig können die Lernsituationen durch unterschiedliche Arten von Mehrsprachigkeit geprägt sein (vgl. Geist 2021: 77–85). In diesem Kontext sind unter Umständen Ähnlichkeiten und Unterschiede der beiden (oder mehr) im Spiel befindlichen Sprachen von Bedeutung (vgl. Meisel 2011: 245). Ein nächster wichtiger Faktor, der sich ebenfalls aus der zeitlichen Abfolge der Erwerbssituationen ergibt, ist das Alter, in dem eine Sprache gelernt wird. Dass das Erwerbsalter eine Rolle dabei spielen kann, wie sprachliche Entwicklungen ablaufen, wurde bereits angedeutet und es gibt eine Reihe von Studien, die sich mit der Frage beschäftigen, welche Rolle der Altersfaktor tatsächlich spielt (vgl. Czinglar 2014; für eine Zusammenfassung vgl. Bordag / Opitz 2021: 207–208). Seit einigen Jahren wird aus unterschiedlichen Perspektiven verstärkt herausgearbeitet, dass der Altersfaktor in diesem Kontext eher auf den Maturitätsgrad des Gehirnes bezogen werden muss (vgl. Meisel 2011: 249; Friederici 2017) und gleichzeitig keine klaren Trennlinien zwischen verschiedenen Erwerbssituationen aufgemacht werden können (vgl. aus neurowissenschaftlicher Perspektive DeLuca et al. 2019). Weitere Faktoren, die den Fremd- und Zweitspracherwerb in unterschiedlicher Weise als die erstsprachliche Entwicklung zu beeinflussen scheinen, sind die Vorbildung und der soziale bzw. sozioökonomische Status von Lernenden (vgl. Czinglar et al. 2015; Grießhaber 2018: 25) sowie die generelle Sprachlernbegabung (*aptitude*; vgl. Hufeisen / Riemer 2010: 745–746).[32] Dazu kommen Persönlichkeitsfaktoren wie

the Native Speaker Put to the Test: Recent Research Advances" (vgl. Guijarro-Fuentes et al. (2021|2022); vgl. für das Deutsche als L1 Shadrova et al. 2021). Aus neurowissenschaftlicher Perspektive erhält die Diskussion um durchaus heterogene (bilinguale) erstsprachliche bzw. mehrsprachliche Entwicklungsverläufe etwa durch die Studie von DeLuca et al. (2019) Unterstützung, die dafür plädieren, die starke Dichotomie zwischen verschiedenen Erwerbstypen aufzulösen und eher von einem Spektrum auszugehen, in dem unterschiedliche Erfahrungen und Voraussetzungen der Lernenden eine wichtige Rolle spielen.

32 Sprachlerneignung oder *language aptitude* bezeichnet „ein kognitives Konstrukt […], das zumeist mit einem spezifischen Test operationalisiert wird (Modern Language Aptitude Test: MLAT), der phonologische Diskriminierungs- und Enkodierungsfähigkeit, grammatisches Einfühlungsvermögen, induktive Sprachlernfähigkeit sowie die Fähigkeit zum Auswendiglernen misst" (Hufeisen / Riemer 2010: 745–746).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Motivation, Angst, „Ambiguitätstoleranz (Umgang mit Widersprüchlichkeiten), Empathie (Einfühlungsvermögen), Risikobereitschaft (insbesondere bei der Sprachproduktion)" (Hufeisen / Riemer 2010: 746) und nicht zuletzt individuelle Lernstile. Zusätzlich sind Lernende beim Lernen einer L2 handelnde Subjekte, die in interaktionalen Beziehungen zu L1-Sprechenden und anderen L2-Sprechenden die L2 lernen (vgl. Grießhaber 128; vgl. Königs 2010: 762). Werden zusätzlich gesteuerte Erwerbskontexte, wie der schulische Deutscherwerb, in die Überlegungen miteinbezogen, spielen neben curricularen und bildungspolitischen Vorgaben vor allem Einflussfaktoren wie „Lehrer- und Schülerpersönlichkeit, Faktoren der Unterrichtsgestaltung, sozialer Zusammenhalt der Klassengemeinschaft, persönliche Freundschaften der Schüler/innen […]" (Kuhs 2020: 528) eine entscheidende Rolle.

Dieser kurze Vergleich zwischen erstsprachlichen Prozessen auf der einen Seite sowie fremd- und zweitsprachlichen Prozessen und den sie beeinflussenden Faktoren auf der anderen Seite macht deutlich, dass Zweitspracherwerbsstudien eine Reihe von miteinander zusammenhängenden und sich gegenseitig beeinflussenden Variablen in den Blick nehmen müssen. Wie dies auf der sprachlichen Oberfläche von lernersprachlichen Äußerungen umgesetzt werden kann, verdeutlicht das nächste Kapitel.

2.2 Erforschung und Analyse von Lernersprache

Die Beschäftigung mit Lernersprache, die an dieser Stelle kurz umrissen werden soll, stellt eine Möglichkeit dar, zweitsprachliche Entwicklungen einerseits nachzuvollziehen und andererseits zu analysieren. Dabei konzentrieren sich die Ausführungen, wie im vorherigen Kapitel, auf wenige Ausschnitte und Ansätze, die für die vorliegende Studie von besonderer Bedeutung sind. Im Kontext der Analyse von Lernersprache sind dies das Interlanguage-Konzept nach Selinker (1972) und die Bewertung von Lernersprache vor dem Hintergrund des CAF-Frameworks (vgl. Skehan 1989; 1996; 1998).

Einen der großen Durchbrüche im Bereich der Zweitspracherwerbsforschung stellt der Interlanguage-Ansatz von Selinker (1972) dar (vgl. Larsen-Freeman 2014: 204–205). Der durch Selinker angestoßene Paradigmenwechsel betrifft beispielsweise die konzeptionelle Trennung von Lehren und Lernen und die Betrachtung von Lernersprachen (vgl. Selinker 1972: 209–210) als eigene und separate linguistische Systeme, die den linguistisch-kognitiven Raum zwischen einer Ausgangs- und einer Zielsprache aufspannen (vgl. Selinker 2014: 223).[33]

33 Selinker macht einen definitorischen Unterschied zwischen Interlanguage und Lernersprache. Ihm geht es vor allem darum, dass die Sprachen von fortgeschrittenen L2 Sprechenden „with complex grammaticalized syntax" (Selinker 2014: 229), nicht mehr Lernersprachen genannt werden können, selbst wenn sie Hinweise für abweichende phonetische, syntaktische und semantische Systeme

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Lernende werden so quasi zu L1-Sprechenden ihrer eigenen Lernersprachen (vgl. Corder 1973). Dazu kommt das sehr relevante Konzept der Fossilisierung, das den Umstand bezeichnet, dass sich bestimmte nicht-zielsprachliche Strukturen etwa im Bereich der Wortstellung in bestimmten Situationen in den Lernersprachen verfestigen und nicht mehr ohne großen Aufwand auflösbar sind. Dieser Prozess des ‚Steckenbleibens' macht für Selinker einen der wesentlichen Unterschiede zwischen Erst- und Zweitspracherwerb aus (vgl. Selinker 2014: 226). Darüber hinaus liefert Selinker (1972) eine erste Definition von (nicht) akzeptablen Daten zur Erforschung zweitsprachlicher Entwicklung. Selinker geht hier davon aus, dass etwa Aufgaben, bei denen Grammatikalitätsurteile getroffen werden, Aufgaben mit Kunstwörtern oder mit einzelnen Silben keinen Aufschluss über tatsächliche Erwerbsprozesse liefern können (vgl. Selinker 1972: 210; vgl. Selinker 2014: 223). Ortega (2014) ergänzt die klare Benennung von drei extrem einflussreichen Konstrukten, die sich durch den Interlanguage-Ansatz in der Zweitspracherwerbsforschung etabliert haben: Ausgangssprache (*Native Language*, NL), Zielsprache (*target language*, TL) und natürlich Interlanguage bzw. Lernersprache (IL) (vgl. Ortega 2014: 175).

Lernersprache wird oft anhand von drei Perspektiven untersucht, die als Ganzes als CAF-Ansatz zusammengefasst werden.[34] CAF steht für *complexity* (Komplexität), *accuracy* (Korrektheit bzw. Angemessenheit) und *fluency* (Flüssigkeit bzw. Automatisierungsgrad) (vgl. Housen / Kuiken / Vedder 2012; Housen / Kuiken / Vedder 2012a).[35] Während die Ursprünge dieses Ansatzes bis in die 1960er-Jahre zurückreichen (vgl. Housen et al. 2012a: 2), werden die drei Dimensionen zum ersten Mal von Skehan (1989; 1996; 1998) in einem integrierten Modell zusammengefasst, wobei die Konstrukte selbst als „vielschichtig, facettenreich und multidimensional" (Housen et al. 2012a: 5; [Übersetzung MS]) aufzufassen sind. Diese drei Dimensionen bilden zusammen ein Set von Möglichkeiten, um die Entwicklung und den Zustand von Lernersprachen mithilfe von quantitativen Variablen beschreiben zu können (vgl. Rosmawati 2020: 111). Sie sind mit-

zeigen (vgl. ebd.). Im Kontext der vorliegenden Arbeit, die zweitsprachliche Entwicklung aus einer gebrauchsbasierten und komplex-dynamischen Perspektive betrachtet, wird durchgängig der Begriff Lernersprache verwendet. Während Interlanguage einen Prozess hin zu einer Zielsprache impliziert, ist der Begriff Lernersprache etwas offener und deutet einen kontinuierlichen, letztlich nicht abschließbaren Prozess an.

34 An dieser Stelle kann im Folgenden nur ein knapper und in keinem Fall erschöpfender Überblick über den CAF-Ansatz, seine Entstehung und die innerhalb dieses Ansatzes operationalisierten Maße gegeben werden. Einführende Überblicksdarstellungen und kritische Reviews der bisher operationalisierten Maße finden sich beispielsweise bei Wolfe-Quintero et al. (1998), Norris / Ortega (2009), Pallotti (2009), Housen et al. (2012a), Bulté / Housen (2012) und in Bezug auf linguistische Komplexität Pallotti (2015).

35 Im Folgenden werden in der Regel die englischen Begriffe zur Beschreibung der drei Dimensionen verwendet.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

einander verschränkt und entwickeln sich in gegenseitigen, zum Teil nichtlinearen Abhängigkeiten (vgl. Housen et al. 2012a: 7; vgl. Spoelman / Verspoor 2010: 532 und vgl. zu möglichen Konzeptualisierungen dieser Beziehungen vor allem das Kapitel 2.5 zur *Complex Dynamic Systems Theory*). Es wird angenommen, dass sich mithilfe der CAF-Dimensionen Rückschlüsse auf die sprachlichen Fertigkeiten von Lernenden ziehen lassen und der Ansatz in Konkurrenz zum „traditional four-skills model" der Fertigkeiten Hören, Lesen, Sprechen und Schreiben steht (Housen et al. 2012a: 1).[36] *Complexity* steht für die Elaboriertheit einer Lernersprache bzw. die Fähigkeit von Lernenden, eine Reihe von unterschiedlichen grammatischen Formen (d. h. Strukturen, Muster oder Regeln; vgl. Housen et al. 2019: 8) und lexikalischen Einheiten zu verwenden (vgl. Housen et al. 2012a: 2). *Accuracy* dient als Maß für die Angemessenheit bzw. die zielsprachliche Verwendung sprachlicher Strukturen (vgl. ebd.) und *fluency* misst den Grad der Automatisierung lernersprachlicher Produktionen (vgl. Rosmawati 2020: 111). Einschränkend ist anzumerken, dass die drei grundlegenden Konstrukte bzw. Dimensionen des Ansatzes oft entweder nicht klar oder zu allgemein definiert werden oder sich die angebotenen Definitionen nicht decken (vgl. Housen et al. 2012a: 3–4). Für die vorliegende Studie spielt vor allem das Konstrukt der Komplexität (*complexity*) eine Rolle, weswegen dieses im Folgenden näher beschrieben werden soll.

In den letzten Jahren wurden zahlreiche Versuche unternommen, Komplexität als Konstrukt in der Spracherwerbsforschung fruchtbar und erklärbar zu machen.[37] Es ist zu beachten, dass es bis heute keine allgemein akzeptierte Definition von Komplexität gibt (vgl. Housen et al. 2019: 4; vgl. Pallotti 2015: 118–119) und dass sich der Begriff selbst durch eine relativ deutlich hervortretende Mehrdeutigkeit auszeichnet. Komplexität wird vielmehr in verschiedener Weise operationalisiert und erhält dadurch kontextabhängig jeweils eine unterschiedliche Bedeutung (vgl. Housen et al. 2019: 4).[38]

Für die Zweitspracherwerbsforschung stellen Bulté / Housen (2012) eine umfassende Taxonomie unterschiedlicher Komplexitätskonstrukte vor (vgl. Bulté / Housen 2012: 23). Die beiden Autoren unterscheiden zunächst relative und absolute Komplexität, wobei relative Komplexität nur individuell bestimmbar

36 Es ist fraglich, ob das CAF-Framework als eigener theoretischer Ansatz beschrieben werden kann, da viele Studien, die in diesem Ansatz verortet sind, sehr heterogene Theorien zugrunde liegen bzw. keine ‚einheitliche' auf den CAF-Framework orientierte Theoriebildung existiert.

37 Vgl. für einen Überblick über Adaptionen in verschiedenen Fachbereichen der Linguistik Housen et al. (2019: 4).

38 Um die definitorische Unschärfe, die aus verschiedenen Operationalisierungen von Komplexität entspringt, abzuschwächen, schlägt Pallotti (2015) vor, diese nur noch als deskriptive Kategorie zu betrachten, die nicht direkt auf Hypothesen über die Entwicklung von Lernenden übertragen werden kann (vgl. Pallotti 2015: 119).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

ist und sich aus der Beziehung von Sprachnutzenden zu bestimmten Strukturen bestimmen lässt. Sie wird daher oft mit Schwierigkeit (*difficulty*) bzw. mit kognitiver Verarbeitbarkeit assoziiert (vgl. ebd.). Die absolute Komplexität als inhärente Eigenschaft der beobachteten Strukturen hingegen wird von Bulté / Housen in drei Teilkonstrukte differenziert: Linguistische Komplexität (*linguistic complexitxy*), eine Komplexität, die sich auf die Propositionen von Texten bzw. Aussagen bezieht (*propositional complexity*) und zuletzt diskursiv-interaktionelle Komplexität (*discourse-interactional complexity*), die sich eher auf dialogische Situationen und etwa auf Sprecherwechsel bezieht (vgl. Bulté / Housen 2012: 23). Für die vorliegende Arbeit ist vor allem die linguistische Komplexität von Bedeutung, die in System- und Strukturkomplexität weiter unterteilbar ist (vgl. ebd.). Dabei steht die Lernersprache für das System und die einzelnen linguistischen Formen, die untersucht werden, für die Strukturebene (vgl. ebd.: 25). Weiterhin kann linguistische Komplexität in grammatische und lexikalische Komplexität unterteilt werden (vgl. ebd.: 26), wobei grammatische Komplexität sich aufspaltet in morphologische Komplexität (vgl. Pallotti 2015: 121–123; vgl. Brezina / Pallotti 2019) und syntaktische Komplexität, wie sie in der vorliegenden Arbeit untersucht wird (vgl. Bulte / Housen 2012: 27). Durch die Fortschritte in der automatisierten Berechnung von Komplexitätsmaßen wie der Länge von Sätzen oder Phrasen oder dem Anteil bestimmter Wortarten innerhalb der genannten Strukturen durch computerlinguistische Werkzeuge wie dem *Automatic Syntactic Complexity Analyzer* (Lu 2010) oder der *Common Text Analysis Plattform* (CTAP, Chen / Meurers 2016) lassen sich mittlerweile eine Vielzahl unterschiedlicher Komplexitätsmaße relativ mühelos aus (lernersprachlichen) Texten extrahieren. Bereits Bulté / Housen (2012: 35) machen auf die große Anzahl verfügbarer Maße aufmerksam. Werkzeuge wie CTAP können mehrere hundert unterschiedliche Maße automatisiert aus Sprachbeispielen berechnen.[39] Komplexitätsmaße werden in vielen Fällen durch unterschiedliche Faktoren beeinflusst und geben daher Aufschluss über mehr als eines der beschriebenen bzw. durch die jeweiligen Maße anvisierten Konstrukte (für eine Diskussion vgl. Bulté / Housen 2012: 35–49; vgl. Kapitel 3.2.5 für eine methodische Diskussion der in der vorliegenden Studie verwendeten Komplexitätsmaße). Eine der mittlerweile durchaus kritisch reflektierten Grundannahmen der Forschung im Bereich linguistischer Komplexität ist die Annahme, dass steigende Werte in verschiedenen Komplexitätsmaßen direkt mit zunehmender Komplexität assoziiert werden können und zunehmende Komplexität wiederum ein direktes Maß für zunehmende Kompetenz der Lernenden darstellt (vgl. Housen et al. 2019: 8).[40] Eine solche Argumentation trägt in besonderer Weise die Gefahr

39 Für das Deutsche stehen derzeit in CTAP 543 verschiedene Maße zur Verfügung (Stand 9.2.2022), die automatisiert aus Texten extrahiert werden können.

40 Petersen (2014) etwa assoziiert den Anstieg der Textlänge bei beginnenden Schreibenden mit einer zunehmenden Automatisierung von Schreibprozessen (vgl. Petersen 2014: 83).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

eines Zirkelargumentes in sich, das folgendermaßen zusammengefasst werden könnte: Komplexe Strukturen sind Strukturen, die spät erworben werden und eine Struktur wird spät erworben, weil sie komplex ist. Dass die Werte von Komplexitätsmaßen im Laufe der zweitsprachlichen Entwicklung zunehmen, stellt keine Validierung dieser Maße dar (vgl. Bulté / Housen 2012: 36). Um Entwicklung umfassend nachzeichnen zu können, müssten sich diese Maße über die gesamte Zeit der zweitsprachlichen Entwicklung erstrecken, was in vielen Fällen methodisch herausfordernd sein könnte, da diese unterschiedlichen Maße im Zeitverlauf in unterschiedlichen Abhängigkeiten und Beziehungen zueinanderstehen und können (vgl. ebd.). Darüber hinaus ist von einer großen Aufgabenabhängigkeit dieser Maße auszugehen.

2.3 Schreiben in Deutsch als Zweitsprache

Die Fähigkeit zu schreiben und die Entwicklung dieser Fähigkeit ist in besonderem Maße in schulischen bzw. institutionellen Bildungskontexten verortet und an die Förderung in diesen geknüpft (vgl. Schmölzer-Eibinger 2018: 3–4; vgl. Byrnes 2020: 245).[41] Dies hängt unter anderem mit dem hohen Normierungsgrad geschriebener Sprache in verschiedenen Sprachbereichen (wie z. B. Lexik, Grammatik, Aussprache und Orthografie zusammen (vgl. Roll (2021: 270). Geschieht das Schreiben zusätzlich in einer Zweitsprache, ist es als Teil des Zweitspracherwerbs darüber hinaus zahlreichen weiteren, individuellen Einflussfaktoren unterworfen (vgl. Kapitel 2.1). Es wird daher als „vielschichtig und heterogen" (Grießhaber 2020: 292) konzeptualisiert und zusätzlich von vielen verschiedenen theoretischen Perspektiven aus in den Blick genommen (vgl. Baba / Nitta 2014: 2).

Dieses Kapitel beschäftigt sich überblicksartig mit wesentlichen Erkenntnissen der Schreibforschung im Kontext des Deutschen als Fremd- und Zweitsprache und bezieht sich auf mögliche Transfereffekte aus der L1 bei der Produktion von schriftlichen Texten sowie die Konstrukte des Schreibprozesses, der Schreib- bzw. Textkompetenz und der Messung von Textqualität. Ebenfalls wird diskutiert, inwiefern sich anhand schriftlicher Produkte die syntaktische Kompetenzentwicklung im Bereich Schreiben rekonstruieren lässt. Da es sich bei einer solchen Darstellung notwendigerweise lediglich um einige kursorische Anmerkungen zu einem sehr komplexen und vielschichtigen Thema handeln kann, werden immer wieder Verweise auf weiterführende Literatur gegeben.

Der Stand der Forschung zum zweitsprachlichen Schriftspracherwerb bzw. zur Entwicklung zweitsprachlicher Schreibkompetenz hat sich vor allem hin-

41 Studien, die das Schreiben bzw. Schreibprodukte untersuchen, sollten dies daher stets vor dem Hintergrund eben dieser institutionellen bzw. schulischen Kontexte tun und diese systematisch in den Forschungsdesigns reflektieren.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

sichtlich des Zweitspracherwerbs in schulischen Kontexten (und hier vor allem in Bezug auf die Grundschule) in den letzten Jahren deutlich verbessert (vgl. hierzu etwa den Überblick in Ricart Brede 2020: 60f.; vgl. Petersen 2014: 42f; vgl. vor allem das Handbuch zum Schreiben in der Zweitsprache Deutsch von Grießhaber et al. 2018). Schramm (2018) spricht in diesem Kontext von einer „rasanten" (Schramm 2018: 33) Entwicklung der Forschung zum zweitsprachlichen Schreiben. Für das sich entwickelnde L2-Schreiben (ab dem Anfangsstadium) von beginnenden, erwachsenen Zweitsprachlernenden ist die Quellenlage deutlich eingeschränkter (vgl. Domenech / Petersen 2018: 117; vgl. Demmig 2018: 411). Insgesamt scheint diese Zielgruppe seltener explizit in den Fokus genommen zu werden. Eine Ausnahme bildet in jüngster Zeit das Projekt „Deutsch im Beruf" des IDS Mannheim und des Goethe-Instituts, welches Geflüchtete in Integrationskursen begleitet. In diesem Projekt werden jedoch nur mündliche Textproduktionen erhoben (vgl. Cindark et al. 2019; Deppermann et al. 2018; Hünlich et al. 2018).[42] Einige Untersuchungen und Forschungsprojekte untersuchen fortgeschrittene Lernende in L2-Kontexten (etwa an Universitäten oder weiterführenden Schulen). In diesen Fällen haben die Lernenden bereits eine umfangreiche Schreib- und Textkompetenz Deutsch als L2 erworben (vgl. hierzu die Darstellungen in Knorr (2018) zur zweitsprachlichen Schreibentwicklung in der Hochschule und Niederhaus (2018) zum zweitsprachlichen Schreiben in der beruflichen Bildung). In jüngster Zeit nimmt das SpraStu-Projekt des Herder-Instituts der Universität Leipzig und der Universität Würzburg fortgeschrittene Lernerinnen und Lerner anhand schriftlicher Daten in den Blick (vgl. Wisniewski et al. 2022; vgl. Wisniewski 2018; vgl. Kapitel 1).

Insgesamt lässt sich für die L2-Schreibforschung eine ähnliche methodische Entwicklung wie für andere Forschungskontexte beobachten (vgl. dazu Kapitel 1). Matias beschreibt diese Entwicklung als „Bewegung von einer qualitativ-explorativen hin zu einer experimentell-deskriptiven Orientierung" (Matias 2017: 62).[43] Es werden folglich mehr Studien durchgeführt, die eher quantitativ ausgerichtet sind oder mehrere Methoden miteinander kombinieren. Zu einem ähnlichen Schluss kommt Schramm (2018), die außerdem darauf aufmerksam macht, dass die deutschsprachige L2-Schreibforschung „in charakteristischer Weise durch das Zusammenspiel von Linguistik einerseits und Lehr- und Lernforschung andererseits beeinflusst ist" (Schramm 2018 34). Obwohl der Schreibprozess selbst nicht direkt Gegenstand der vorliegenden Studie ist, sollen einige für die (L2-) Schreibforschung relevante Schreibprozessmodelle überblicksartig in den Blick

42 Ricart Brede (2020:61) führt zahlreiche Studien an, die die Alphabetisierung erwachsener L2-Lernender fokussieren und die daher für den Kontext dieser Arbeit nicht weiter relevant sind.

43 Selbst wenn Matias (2017) eher Forschungsarbeiten im Blick hat, die sich auf den Schreibprozess selbst und nicht die Produkte dieses Prozesses konzentrieren, scheint seine Feststellung für die hier genannten Forschungsarbeiten in ähnlichem Maße zu gelten.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

genommen werden, um grundsätzliche Erwerbsaufgaben der Lernenden verstehen zu können.[44] Modelle des Schreibprozesses „versuchen, die Gesamtheit der komplexen Vorgänge, die während der Produktion eines Textes bei einem Individuum ablaufen, abzubilden" (Petersen 2014: 36). Für das Deutsche als L2 ist hier vor allem das durch Grießhaber (2016: 220; 2020: 299) adaptierte Modell von Hayes / Flower (1980) relevant.[45] Dieses nimmt drei spezialisierte „Werkstätten" (Grießhaber 2020: 299) an, die die unterschiedlichen Aspekte des Schreibprozesses steuern. Diese sind die Aufgaben-Werkstatt, die Wissens-Werkstatt und die Prozess-Werkstatt. Alle diese Werkstätten zeichnen sich außer durch ihre spezifischen Aufgabenbereiche dadurch aus, dass sie bestimmten Einflüssen der L2 unterliegen (vgl. ebd.). Die Aufgaben-Werkstatt bezieht sich auf das Aufgabenumfeld der konkreten Schreibaufgabe, die bearbeitet wird. Zusätzlich können hier motivationale Aspekte und die angenommene Rezeption der Textproduktion eine Rolle spielen (vgl. ebd.). Die Wissens-Werkstatt wird mit dem Langzeitgedächtnis assoziiert und ist vor allem dafür verantwortlich, Schreibpläne zu speichern. Die dritte Werkstatt bildet als Prozess-Werkstatt den Kern des Schreibprozesses. Sie ist auf der Mikro- und Makroebene verantwortlich für die Planung und Organisierung des zu verfassenden Textes (vgl. ebd.). Für die Prozess-Werkstatt nimmt Grießhaber die stärksten L2-Einflüsse an, „da hier Wortschatz und Grammatik der verschiedenen Sprachen eine zentrale Rolle zukommt" (Grießhaber 2020: 299). Während das Modell den großen Vorteil hat, dass es den gesamten Schreibprozess in kleinere, didaktisch verarbeitbare Einheiten aufteilt, werden bestimmte relevante Faktoren wie Alter, Schreiberfahrung und motivationale Aspekte ausgeklammert (vgl. Roll 2021: 274).

Grundsätzlich stellt sich die Frage, inwiefern die bisher eher für Kinder und Jugendliche herausgearbeiteten Entwicklungen auf erwachsene L2-Schreiber/-innen übertragbar sind bzw. durch was diese spezifisch gekennzeichnet sind.[46] Ricart Brede stellt die großen Unterschiede zwischen L2-Lernenden als „generelle Schwierigkeit für die Erforschung des Schreibens in einer L2" (Ricart Brede

44 Für eine ausführliche Zusammenfassung des Forschungsstandes zu Schreibprozessen vgl. Petersen (2014: 36–40), für einen methodischen Überblick im Kontext der Schreibprozessforschung vgl. Matias (2017). Ebenso finden sich methodische Überblicksdarstellungen zur Schreibprozessforschung in Linnemann (2017) sowie Weinzierl / Wrobel (2017). Ein weiterer Überblick, der die gesamte DaZ-Schreibforschung in den Blick nimmt, findet sich bei Schramm (2018).

45 Erklärungen und die grafische Repräsentation des von Grießhaber adaptierten Modells finden sich jeweils bereits in den ersten Auflagen der hier genannten Beiträge bzw. Monografien, außerdem wird das Modell ausführlich in Petersen (2014: 40–41) besprochen. Die folgenden Ausführungen beziehen sich vor allem auf Grießhaber (2020), dessen Beitrag in der fünften und unveränderten Auflage von „Deutsch als Zweitsprache" (Ahrenholz / Oomen-Welke 2020) erschienen ist.

46 Zusammenfassende Darstellungen wesentlicher Erkenntnisse zur zweitsprachlichen Schreibentwicklung bei Kindern und Jugendlichen finden sich bei Jeuk (2018) und Becker (2018).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

2020: 61) fest und hebt hier vor allem unterschiedliche Erstsprachen, das Alter zu Beginn des Zweitspracherwerbs und die bereits gesammelten Schreiberfahrungen in der L1 sowie die erreichte Kompetenz in der L2 hervor (vgl. ebd.).[47] Herausforderungen für L2-Schreibende können sich daher in Differenz zu ihrer L1 ergeben, etwa, wenn die Phonem-Graphem-Korrespondenz zwischen L1 und L2 stark variiert oder falls, wie im Fall des Sprachenpaares Arabisch-Deutsch, die Schreib- und Leserichtung differiert oder ein anderes Alphabet zugrunde liegt (vgl. Ricart Brede 2020: 64f.; vgl. Roll 2021: 270).[48] Weiter führt Ricart Brede aus, dass „die individuellen Schreibbiographien und die schreibrelevanten Vorerfahrungen (im Deutschen als L2 und in anderen Sprachen) umso stärker an Bedeutung [gewinnen], je älter die Schreibenden sind" (ebd.).[49] Gleichzeitig gehen viele Autorinnen und Autoren von bestimmten Wechselwirkungen zwischen L1 und L2 im Bereich von „Textproduktionskompetenz" (Ricart Brede 2020: 66) aus.[50] Literale Strukturen aus der L1 wirken sich auf die L2 aus und umgekehrt (vgl. Roll 2021: 272). Dabei werden bestimmte Erkenntnisse als relativ gesichert gesehen (vgl. hierzu Grießhaber 2020: 293; Ricart Brede 2020:

47 Ricart Brede spricht hier von einer „kritische[n] Schwelle" (Ricart Brede 2020: 72). Sie geht davon aus, dass das Überschreiten einer solchen Schwelle die Voraussetzung für bestimmte Transfers aus der L1 bildet. Da die Teilnehmenden der vorliegenden Studie alle angenommenen kritischen Schwellen bereits überschritten haben (vgl. hierzu Kapitel 3.1.2), wird an dieser Stelle nicht weiter auf die Diskussion um kritische Schwellen und Perioden eingegangen.

48 Auf einen systematischen und umfassenden Vergleich des Sprachpaares Arabisch-Deutsch wird an dieser Stelle und in dieser Arbeit insgesamt verzichtet, da die aufgeworfenen Fragestellungen nicht aus einer kontrastiven Perspektive gestellt wurden, siehe aber Kapitel 1 und Kapitel 3.1.2 sowie die Diskussion am Ende dieser Arbeit für weitere Hinweise und weiterführende Literatur zum kontrastiven Vergleich Arabisch-Deutsch und zu diesem Thema aus Lernendenperspektive.

49 In der vorliegenden Studie wurde versucht, viele dieser Faktoren zu kontrollieren (vgl. dazu Kapitel 3.1.1). Trotz allem ist von einer insgesamt sehr heterogenen Gruppe von Teilnehmenden auszugehen (vgl. dazu Kapitel 3.1.2). Die Daten der hier vorliegenden Studie entstammen Schreibproduktionen von Anfängerinnen und Anfängern, ab dem Beginn ihrer sich entwickelnden L2. Datensets, die in vergleichbarer Weise Anfänger/-innen in den Blick nehmen, finden sich etwa bei Vyatkina (2012; 2013; 2015), die mit den schriftlichen Daten von beginnenden Lernenden aus den USA das longitudinale *Kansas Developmental Learner Corpus* (KanDeL; vgl. Vyatkina 2016) erstellt hat. Das KanDeL-Korpus ist Teil der Falko-Korpus Familie der Humboldt-Universität zu Berlin und über die ANNIS-Suchmaske zu durchsuchen. Weitere Informationen finden sich unter: https://www.linguistik.hu-berlin.de/en/institut-en/professuren-en/korpuslinguistik/research/kandel [23.02.2023].

50 Roll (2021) spricht in diesem Zusammenhang von Textkompetenz, die „als sprachlich-kognitive Fähigkeit verstanden [wird], Texte selbständig oder in Kooperation mit anderen zu produzieren, zu verstehen und als Grundlage für das weitere Denken, Sprechen und Handeln zu nutzen" (Roll 2021: 270).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

66f.). Dazu gehört, dass Schreibkompetenzen aus der L1 in der L2 verfügbar sein können. Die Entwicklung von Schriftlichkeit in einer L2 beruht auf Wissen „aus früherer Aneignung von Schriftlichkeit" (Montanari 2018: 188) aus einer L1. Ebenso können bereits bekannte textuelle Planungs- und Überarbeitungsstrategien in der L2 genutzt werden. Roll (2021) nennt hier unter anderem die Fähigkeiten, Kohärenz herzustellen, Sachverhalte zu kontextualisieren und verschiedene Perspektiven darzustellen (vgl. Roll 2021: 272). Textmusterkompetenz scheint grundsätzlich von einer L1 in eine L2 übertragbar, jedoch aufgrund bestimmter „Unterschiede in Bezug auf die Gestaltung bzw. das Vorkommen bestimmter Textsorten teilweise […] einzelsprachlich gebunden" (Ricart Brede 2020:67) zu sein. Das bedeutet, so Ricard Brede weiter, dass es unter Umständen zu „negativen Transfereffekten [und] Interferenzen" (ebd.) kommen kann. Lexikalische und grammatische Mittel der L2 auf der Formulierungsebene hingegen können L2-Lernenden fehlen, da diese in größerem Maße einzelsprachspezifisch geprägt sind. Dies kann zu einem geringeren Textumfang und einer weniger stark ausgeprägten syntaktischen Integration der produzierten Texte führen (vgl. Grießhaber 2014: 3). Zum Teil werden fehlende schriftsprachliche Mittel durch die Verwendung mündlicher Sprache kompensiert.

Wie genau die Textqualität, die als Annäherung an das mental repräsentierte Konstrukt der Schreib- bzw. der Textkompetenz betrachtet wird, messbar gemacht werden kann, ist gerade für erwachsene Lernende nicht so ausführlich erforscht wie für Kinder und Jugendliche (vgl. Feick 2018: 250).[51] In Bezug auf die Produktion schriftlicher Texte ist bis heute nicht abschließend geklärt, wie syntaktische Komplexität und die Qualität von Schreibprodukten zusammenhängen (vgl. Ortega 2015: 87).[52] Wie in dem im vorherigen Kapitel besprochenen CAF-Ansatz werden daher „systematisch sprachliche Phänomene in Texten erfasst, von denen angenommen wird, dass sie für die Produktion von Texten relevant sind und so versucht, die Qualität von Texten abzubilden" (Petersen 2019: 14). Dabei können unterschiedliche Phänomene und Strukturen für die Beurteilung von Texten in Frage kommen. Dies sind wie im CAF-Ansatz beispielsweise quantifizierbare, linguistische Oberflächenmerkmale von Texten wie die Länge bestimmter Strukturen oder die Verwendung von subordinierenden Neben- oder Relativsatzstrukturen (vgl. Petersen 2019: 14; vgl. Kapitel 2.2).

51 Feick merkt zusätzlich an, dass derzeit zwar „Modelle für den zweitsprachlichen Schreibprozess [existieren], aber kein Schreibkompetenz(stufen)modell, das die Basis für die Entwicklung förderdiagnostischer Verfahren bildet" (Feick 2018: 250).

52 Viele Befunde deuten darauf hin, dass ‚gutes Schreiben' ein sehr vielschichtiges Konstrukt ist und sowohl stark genreabhängig geprägt als auch von dem angenommenen oder tatsächlichen Publikum abhängt (vgl. Ortega 2015: 87). Ortega (2015) macht darauf aufmerksam, dass die Analyse der Entwicklung syntaktischer Komplexität konzeptionell unter anderem mit den Konstrukten der durch Unterricht gesteuerten Aneignung, der Modalität und der Genre- bzw. Taskabhängigkeit der gestellten Aufgabe verbunden werden müsse (vgl. Ortega 2015: 91).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

In den nächsten beiden Kapiteln werden zwei Theorieansätze besprochen, die zweitsprachliche Entwicklung unterschiedlich konzeptualisieren. Dies sind die *Processability Theory* (vgl. Kapitel 2.4) und die *Complex Dynamic Systems Theory* (vgl. Kapitel 2.5). Vor allem im Unterkapitel zur Profilanalyse nach Grießhaber (vgl. Kapitel 2.4.3) werden Instrumente zur Diagnose und zur Messung zweitsprachlicher Schreibentwicklung und syntaktischer Komplexität aus dem deutschsprachigen Forschungskontext vorgestellt.

2.4 *Processability Theory* und Profilanalyse

Eine der prominentesten Fragen der Zweitspracherwerbforschung ist die Suche nach aufeinanderfolgenden und möglicherweise universalen Erwerbssequenzen bzw. Erwerbsstufen in der zweitsprachlichen Entwicklung von Lernenden, die sich unabhängig von der jeweiligen L1, dem Erwerbskontext und anderen externen Faktoren wie der Motivation oder dem Alter der Lernenden entwickeln (vgl. Hulstijn / Ellis / Eskildsen 2015: 1). Die Ursprünge dieser Theoriebildung gehen vor allem auf die Arbeiten von Dulay / Burt (1973; 1974) und auf Bailey / Madden / Krashen (1974) zurück, die jeweils eine einzige mögliche Erwerbsreihenfolge im Bereich der Syntax des Englischen als L2 postulierten (vgl. Hulstijn / Ellis / Eskildsen 2015: 2). Gerade im Bereich der syntaktischen Entwicklung von L2-Lernenden gelten heute bestimmte sprachübergreifende und an erstsprachliche Entwicklungen angelehnte Erwerbssequenzen als weitgehend gesichert, auf die etwa die L1 oder unterrichtliche Instruktion nur einen begrenzten Einfluss zu haben scheinen (vgl. Ortega 2009; vgl. Meisel 2011: 86). Gleichzeitig bleiben bis heute Fragen der Universalität und der Replizierbarkeit der gefundenen Sequenzen offen (vgl. Hulstijn et al. 2015: 3). In den folgenden Kapiteln werden zunächst die Grundannahmen und die theoretischen Hintergründe von zwei Theoriesträngen, die sich mit Erwerbssequenzen auseinandersetzen, in den Blick genommen. Dies sind die *Processability Theory* (PT) und der funktional-pragmatische Ansatz der Profilanalyse nach Grießhaber (2005; 2006; 2012; 2016; u.a.) speziell für das Deutsche als Fremd- und Zweitsprache. Darauf aufbauend werden die empirisch-methodischen Entwicklungen dieser Ansätze aufgezeigt, um im Anschluss daran auf die Grenzen der Ansätze und auf die an ihnen geäußerte Kritik einzugehen. Dieser Teil schließt mit einem Abschnitt zur Bedeutung der vorgestellten Ansätze für den Fachkontext DaF/DaZ und einer kurz gehaltenen Zusammenfassung der wichtigsten Punkte.

2.4.1 Grundannahmen und theoretischer Hintergrund

Während verschiedene Erwerbsstufenmodelle im Bereich der Zweitspracherwerbsforschung existieren, hat sich vor allem die *Processability Theory* (PT) in den letzten 25 Jahren zu einem überaus wirkmächtigen Erklärungsansatz in der internationalen SLA-Forschung sowie im deutschsprachigen Forschungs-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

kontext DaF/DaZ entwickelt.[53] Die PT und andere Ansätze, die Erwerbsstufen in der Entwicklung einer L2 annehmen, haben gemein, dass sie die zweitsprachliche Entwicklung als einen hierarchisch organisierten, letztlich linear auf eine zu erreichende L1-Norm ausgerichteten und in seinen Grundzügen vorhersagbaren Prozess konzeptualisieren. Die Entwicklung der PT basiert im Wesentlichen auf den angesprochenen Arbeiten der ZISA-Gruppe (vgl. Meisel et al. 1981; vgl. Clahsen et al. 1983).[54] Die ZISA-Forschungsgruppe um Clahsen, Meisel und Pienemann stellt 1983 eine umfangreiche Querschnittsstudie auf der Basis der Sprachdaten von insgesamt 45 Teilnehmenden mit den Erstsprachen Italienisch, Spanisch und Portugiesisch vor, die in einem ungesteuerten Kontext Deutsch als Zweitsprache erwerben.[55] Die Autoren erheben mündliche Sprachdaten mithilfe von Interviews, die verschiedene Lebensbereiche der teilnehmenden Personen abdecken. Die Interviews wurden transkribiert und aus den Transkripten wurden von allen Teilnehmenden jeweils etwa 50 Äußerungen ausgewählt, die analysiert wurden. Clahsen et al. (1983) legen ein Emergenzkriterium für den Erwerb einer Struktur von mindestens fünf Realisierungen in den jeweiligen obligatorischen Kontexten an (ebd.: 96). Das Konzept der Emergenz bezeichnet das erste produktive Auftreten einer bestimmten Struktur in einem lernersprachlichen Text bzw. einer lernersprachlichen Äußerung, wobei unter produktiv verstanden wird, dass es sich bei der identifizierten Struktur nicht um einen Chunk oder eine als Ganzes gelernte Struktur handelt (vgl. Clahsen et al. 1983; Pienemann 1998; für eine ausführliche Diskussion vgl. Pallotti 2007). Obligatorische Kontexte bezeichnen linguistische Kontexte innerhalb von Strukturen wie Sätzen, minimalen satzwertigen Einheiten oder Phrasen, in denen bestimmte Elemente zwangsläufig auftreten müssen. So müssen in bestimmten Fällen Klammerstrukturen realisiert werden, wenn trennbare Verben oder Perfektkonstruktionen verwendet werden. Ebenso liegt ein obligatorischer Inversionskontext vor, wenn das Vorfeld mit einem anderen Element als dem Subjekt besetzt wird. Obligatorische Kontexte für Nebensätze mit Verbendstellung ergeben sich etwa durch Relativpronomen oder unterordnende Konjunktionen. Um darüber hinaus in ihrer Querschnittsstudie die angenommene Sequenz der Erwerbsstufen herausarbeiten zu können, verwenden die Autoren eine Implikationsskala (ausführliche Erklärungen, wie Implikationsskalen konzipiert sind und wie sie gelesen werden können, finden sich in Pienemann (2005a) und in Jansen (2008)).[56] Dieser Art von Analyse liegt

53 Ausführliche Darstellungen der *Processability Theory* finden sich bei Pienemann (1998) und Pienemann (2005a).

54 Eine ausführliche Darstellung der ZISA-Studie von Clahsen et al. (1983) findet sich in Bryant / Rinker (2021: 286–297).

55 Darüber hinaus existieren aus dem ZISA-Projekt longitudinale Daten. Eine Veröffentlichung dieser Daten wurde von der Forschungsgruppe zwar mehrfach angedeutet (vgl. Meisel et al. 1981: 133; vgl. Clahsen et al. 1983: 157), aber letztlich wohl nicht umgesetzt (vgl. Bryant / Rinker 2021: 287).

56 Bei Implikationsskalen handelt es sich um tabellarische Operationalisierungen von Zuständen (in Querschnittsstudien) oder von Entwicklungen (in longitudi-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

der Gedanke zugrunde, dass Erwerbsstufen nur implikationell durchlaufen werden können, d.h., sobald eine neue Struktur bzw. die ihr zugrundeliegende Regel erworben wird, müssen alle darunterliegenden Stufen bereits erworben sein. Lernende können Stufen nicht überspringen. In einer Implikationsskala muss sich folglich zeigen, dass für alle Lernenden, für die beispielsweise Belege für eine Stufen vorliegen, ebenfalls Belege für die Stufen–1 oder Stufen-x rekonstruierbar sind. Die Autoren leiten aus den von ihnen untersuchten Sprachdaten zunächst eine Mehrkonstituentenphase und dann eine Phase ab, in der Äußerungen mit der ‚kanonischen' Wortstellung SVO (Subjekt, Verb, Objekt, für eine genauere Erklärung vgl. Kapitel 1) realisiert werden. Auf diese Stufe folgt eine Phase, in der Adverbiale der SVO-Struktur vorangestellt werden. Anschließend beginnen Lernende damit, Verbteile getrennt voneinander zu realisieren. Daran anschließend können Lernende eine Subjekt-Verb-Inversion durchführen. Darauffolgend nehmen die Autoren eine Stufe an, in der Adverbien in die Verbalphrase integriert werden können und folglich rechts des finiten Verbs auftreten. Eine letzte Phase, die aus den ZISA-Daten rekonstruiert wird, ist die Phase der Verbendstellung in Nebensätzen. Die Ergebnisse der ZISA-Studie legen für diese Erwerbsstufen einen implikationellen und sequenziellen (d.h. die Stufen werden nur in der herausgearbeiteten Reihenfolge nacheinander durchlaufen) Zusammenhang nahe. Methodisch markierte die ZISA-Studie einen bedeutenden Perspektivwechsel in der Zweitspracherwerbsforschung, da systematisch die interindividuelle Variation zwischen den Lernenden in die Analyse der lernersprachlichen Daten einbezogen wurde und die zweitsprachliche Entwicklung konsequenterweise als ein höchst individuell geprägter Prozess betrachtet wurde (vgl. Lenzing 2015: 90). Gleich-

nalen Studien) im Zweitspracherwerb. Dabei werden in den vertikalen Spalten die verschiedenen, von der PT vorausgesagten Erwerbsstufen und in den horizontalen Zeilen entweder die unterschiedlichen Lernenden oder die Daten der einzelnen Zeitpunkte der Studie gezeigt. Dabei wird entweder in die oberste oder in die unterste Spalte die niedrigste Stufe in die Tabelle eingetragen. In einer Querschnittsstudie werden die Lernenden, die die wenigsten Erwerbsstufen erreicht haben, links in der Implikationsskala angeordnet, so soll die Anordnung „the predicted successive acquisition of the stages through time" (Jansen 2008: 204) simulieren. In der Regel signalisiert eine (fettgedruckte) Linie das Emergenzkriterium von einer korrekten Realisierung in mindestens vier obligatorischen Kontexten (in PT-Studien). Für alle Lernenden, die rechts von dieser Linie angeordnet sind, wird angenommen, dass sie die betreffende Stufe erworben haben. Für die Lernenden links der Linie wird entweder angenommen, dass sie eine bestimmte Struktur noch nicht erworben haben oder in den vorhandenen Daten konnten nicht genug Belege für einen Erwerb dieser Stufe gefunden werden (vgl. Jansen 2008: ebd.). Anzumerken ist, dass Implikationsskalen in durchaus unterschiedlicher Form auftreten können und dass die einzelnen Teile der Skala leicht verändert auftreten können (vgl. etwa die Implikationsskalen in Lenzing 2015). Lenzing (2015) behandelt außerdem die Frage, wie hoch die Skalierbarkeit von Implikationsskalen sein muss, damit diese als gültig gelten können und legt für PT-basierte Studien einen Schwellenwert von mindestens 90 % an.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

zeitig wurde das Emergenzkriterium als Erwerbskriterium eingeführt, das damit zum ersten Mal einen tiefgreifenden Wechsel der Perspektive von defizitär orientierten Fehleranalysen hin zu einer Analyse der Lernersprache als eigenes sprachliches System im Sinne einer *Interlanguage* nach Selinker (1972) ermöglichte (vgl. Lenzing / Nicholas / Roos 2019a: 3). Hier wird deutlich, dass das Emergenzkriterium zunächst für mündliche und spontansprachliche Kontexte operationalisiert wurde und daher nicht direkt auf schriftliche Äußerungen von Lernenden übertragen werden kann, die einem anderen Planungsprozess unterliegen (vgl. Nicholas / Lenzing / Ross 2019: 394; Anmerkungen dazu, wie das Emergenzkriterium in der vorliegenden Studie konzipiert ist, finden sich in Kapitel 3.2.4).

Während die Forscher der ZISA-Gruppe sich zunächst noch auf eine „kontextfreie Variante der generativen Phrasenstrukturgrammatik auf der Grundlage von Chomsky (1965)" (Grießhaber 2012: 177) stützen, bezieht sich Pienemann (1998) in der Ausformulierung der von ihm angenommenen Sequenzen innerhalb der *Processability Theory* auf die *Lexical Functional Grammar* (LFG) und bezieht sich hier im Wesentlichen auf die Arbeit von Kaplan / Bresnan (1982). Zur LFG vgl. u. a. Bresnan (2001) und den Handbuchartikel von Butt / King (2015). Die LFG nimmt syntaktische Strukturen mithilfe von drei unterschiedlichen Repräsentationen in den Blick. Die erste dieser Repräsentationen ist die c-Struktur (*constituent structure*, Pienemann / Lenzing (2015: 167) bezeichnen diese Struktur als Phrasenstruktur), die anhand eines Baumdiagramms Hierarchieverhältnisse innerhalb einer Äußerung sichtbar macht und weiterhin die f-Struktur (*functional structure*), „die grammatische Relationen und andere syntaktische Merkmale in einer Attribut-Wert-Matrix kodiert" (Butt / King 2015: 840; [Übersetzung MS]). Eine Erweiterung der ursprünglichen Konzeption der LFG stellt die a-Struktur (*argument structure*) dar, die sowohl semantische als auch syntaktische Informationen beinhaltet und zentrale Rollen wie Agens oder Patiens etc. innerhalb eines Satzes abbildet (vgl. Lenzing 2015: 96). Ein Ziel der LFG ist, linguistische Erkenntnisse so zu codieren, dass sie effektiv berechenbar und in computergestützten Anwendungen implementierbar werden (vgl. Butt / King 2015: 840).

Die PT ist damit eine Theorie, die auf Grundlage eines psychologischen Modells von Verarbeitbarkeit (*processability*) von linguistischen Strukturen (vgl. Levelt 1989) Vorhersagen darüber trifft, welche Strukturen Lernende wann in ihrer zweitsprachlichen Entwicklung beherrschen (vgl. Pienemann 1998: xv) und wie sich die Fähigkeit, in einer L2 zu sprechen, weiterentwickelt (vgl. Jansen 2008: 188). Die PT versucht, wie die Zweitspracherwerbsforschung insgesamt, Antworten auf die Fragen zu finden, warum sich unterschiedliche Lernende ähnlich (oder unterschiedlich) entwickeln (*developmental problem*) und warum Lernende Sprachwissen erwerben können, das unter Umständen nicht im angebotenen Input explizit gemacht wurde (*logical problem*) (vgl. zu den beiden Begriffen Pienemann / Lenzing 2015: 161). Bei der Entwicklung ihrer zweitsprachlichen Fähigkeiten werden Lernende laut Pienemann durch die Architektur der menschlichen Sprachverarbeitung eingeschränkt, die deshalb den

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Ausgangspunkt der PT bildet (vgl. Pienemann 1998a: 1). Die PT macht dementsprechend Vorhersagen über die Entwicklung von Lernenden auf der Basis der psychologischen Verarbeitbarkeit von Strukturen, die empirisch überprüft werden können. Ziel ist, generelle Sequenzen in der Lernersprache aufzudecken, um „Grammatiken, die zu unterschiedlichen Zeitpunkten verarbeitbar sind" (ebd.: 1; [Übersetzung MS]), rekonstruieren zu können, wobei die PT grundsätzlich davon ausgeht, dass die Erwerbssequenzen im Zweitspracherwerb den Sequenzen im Erstspracherwerb ähneln (vgl. Jansen 2008: 189). Pienemann nimmt im Anschluss an Levelt an, dass L1-Sprechenden und L2-Lernenden eine Art aufgabenspezifischer, grammatikalischer Prozessor im Arbeitsgedächtnis (*working memory*, vgl. Pienemann 2005a: 6) zur Verfügung steht, der grammatikalische Informationen kodieren und temporär speichern kann. Je nachdem, welche Strukturen bzw. diesen Strukturen zugrundeliegenden Regeln dieser Prozessor mithilfe von Verarbeitungsprozeduren verarbeiten kann, stehen unterschiedliche Strukturen in den lernersprachlichen Grammatiken zur Verfügung (vgl. Pienemann 1998: 4) oder anders formuliert: In jeder Phase der zweitsprachlichen Entwicklung können Lernende genau jene Strukturen verstehen und produzieren, die durch den angesprochenen Prozessor verarbeitet werden können (vgl. Pienemann / Lenzing 2015: 159). Die erworbenen Strukturen bzw. die ihnen zugrundeliegenden Regeln eröffnen gleichzeitig einen Raum, in dem sich eine Vielzahl von unterschiedlichen Lernergrammatiken zeigen kann und der durch ein hohes Maß an Variation gekennzeichnet ist. Dieser Hypothesenraum (*hypothesis space*, eine sehr ausführliche Beschreibung findet sich bei Pienemann 1998: 231–330) ist weiter den Begrenzungen der bisher erworbenen Verarbeitungsprozeduren unterworfen (vgl. Lenzing 2015: 92). Dabei fügen Lernende Wörter ihrem Lexikon hinzu und erwerben zunächst weniger komplexe Strukturen auf Ebene der Lemmata. In einem nächsten Schritt werden grammatikalische Informationen auf Phrasenebene und zuletzt über komplexere Strukturen hinweg ausgetauscht. Dieser Informationsaustausch kann sich phrasenübergreifend oder phrasenextern auf größere syntaktische Struktureinheiten beziehen (vgl. Pienemann 2005a: 13; Pienemann / Lenzing 2015: 165). Ein Beispiel für eine Informationsübertragung auf Phrasenebene wäre die Angleichung der Formen in einer Nominalphrase wie *eine **schöne Stadt*** (TN079, Zeitpunkt 1, Einladung nach Leipzig).[57] Ein Informationsaustausch über Phrasengrenzen hinweg ermöglicht Lernenden etwa den Bezug von (durch eine weitere Phrase getrennten) Verbteilen aufeinander: *Wir **können***

57 Alle Belege, die hier und im Folgenden zur Erklärung der Erwerbsstufen nach Pienemann angeführt werden, sind authentische lernersprachliche Belege, die dem dieser Arbeit zugrundeliegenden MPI-HI-Korpus entstammen, das in Kapitel 3 genauer vorgestellt wird. In allen Beispielen wird die von den Lernenden realisierte Orthografie unverändert übernommen. In Klammern sind jeweils die Teilnehmendennummer, der Zeitpunkt und der Titel der Arbeitsaufgabe angegeben. Die jeweiligen Rohdaten zu den einzelnen Texten finden sich im Anhang.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

einen Ausflug im Ostsee für zwei Tagen zusammen ***machen***. (TN048, Zeitpunkt 1, Einladung nach Leipzig). Dieser Prozess wird als *feature unification* (vgl. Pienemann 1998: 97f.; Jansen 2008: 189) bezeichnet. Pienemann / Lenzing 2015 beschreiben diesen Vorgang etwas weniger technisch als Informationsanpassung (*information matching*, ebd.: 162). Dieser Prozess ist aus Sicht der PT letztlich dafür verantwortlich, dass die einzelnen Teile eines Satzes zueinander passen (vgl. Pienemann 2005a: 15). Hinsichtlich der syntaktischen Wortstellungsregeln nimmt die PT in der zweitsprachlichen Entwicklung für das Deutsche als L2 ein Modell hierarchisch angeordneter Erwerbsstufen an (vgl. Pienemann 1998, 2005a).[58] Dieses Modell erweist sich als äußert robust und wird bis heute in zahlreichen Erwerbsstudien zum Deutschen als L2 bestätigt (vgl. Ortega 2014: 176–177; vgl. zudem die Ausführungen zu den empirischen Studien in Kapitel 2.4.2). Folgende Stufen werden angenommen:

- *Stufe 1 – Einzelwörter / Zugriff auf einzelne Lemmata*
- Stufe 2 / Stage x– SVO: Kanonische Wortstellung
- Stufe 3 / Stage x+1: ADV: Initiale Position / Adverbiale Voranstellung
- Stufe 4 / Stage x+2: PART / SEP: Trennbare Verben, zwei getrennte Verbteile
- (Stufe 5) / Stage x+3: INV: Subjekt-Verb-Inversion
- (Stufe 6) / Stage x+4: V-END: Verbendstellung in Nebensätzen

Die Stufen zeichnen sich durch jeweils eigene zugrundeliegende Regeln aus, die unterschiedliche Verarbeitungsprozesse bedingen (vgl. Pienemann / Lenzing 2015: 160). Die genannten Stufen sollen im Folgenden kurz skizziert werden. Zu jeder Stufe werden die dazugehörigen Regeln paraphrasiert, wobei die (eher technischen) Formalisierungen in Pienemann (1998) oder Pienemann (2005a: 30–33) zu finden sind. Aus Gründen der besseren Lesbarkeit wurde an dieser Stelle auf die Angabe von Strukturformeln verzichtet, die in den beiden zuletzt genannten Veröffentlichungen von Pienemann aufgeführt werden.

58 Die PT nimmt ursprünglich sechs Erwerbsstufen an, wobei die erste Stufe sich auf den Zugriff auf Lemmata bezieht und daher für die Entwicklung von Wortstellungsregeln im Kontext der vorliegenden Studie keine Rolle spielt. Diese erste Stufe ist daher kursiv gesetzt. Die durch die PT vorausgesagten Erwerbsstufen ähneln weitgehend den Stufen, die aus den ZISA-Daten rekonstruiert wurden, sind jedoch nicht völlig identisch und wurden durch Pienemann et al. (2005) noch um eine weitere Stufe ergänzt. Eine methodische Erweiterung erfuhr die PT durch die TOPIC-Hypothese von Pienemann / Di Biase / Kawaguchi (2005). Die Autoren nehmen an, dass Lernende zunächst nicht zwischen TOPIC und SUBJ differenzieren können. Oder anders gesagt, Lernende ordnen der satzinitialen Position (TOPIC) stets die Funktion Subjekt (SUBJ) zu. Sobald Lernende die satzinitiale Position bzw. im Deutschen das Vorfeld mit etwas anderem als einem Subjekt besetzen und dieses Element der ‚kanonischen' Wortstellung voranstellen, beginnt ein Differenzierungsprozess, der dafür sorgt, dass die Strukturen TOPIC und SUBJ voneinander unterschieden werden (vgl. Pienemann / Di Biase / Kawaguchi 2005: 239; Pieneman / Lenzing 2015: 169f.).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

(Stufe 2) SVO: Canonical word order

Ich habe in den deutschkurs urlaub (TN022, Zeitpunkt 1, Einladung nach Leipzig)

Die erste im Kontext dieser Arbeit genauer zu betrachtende Erwerbsstufe der PT ist Stufe 2, die den Erwerb einer ‚kanonischen' Wortstellung annimmt. In der PT wird die Hypothese aufgestellt, dass Lernende mit einer SVO-Sprache als L1 ebenfalls mit der später zu korrigierenden Anfangshypothese über das Deutsche als SVO-Sprache beginnen und in der Lage sind, diese Wortstellung zu produzieren (vgl. Pienemann 1998, 2005a). Eine SVO-Struktur scheint für Lernende leichter nachzuvollziehen, da die verschiedenen semantischen Rollen so auftreten, dass sie direkt der Position in der unterliegenden c-Struktur entsprechen (vgl. Jansen 2008: 191). Sie entsprechen damit, mit anderen Worten, der Reihenfolge, die am unmittelbarsten zugänglich ist, da in der Regel in der ersten Nominalphrase das handelnde Subjekt bzw. das Agens des Satzes genannt wird und dann beschrieben wird, was dieses tut. Unterschiedliche empirische Studien haben abweichende Muster beim Erwerb der Wortstellung ergeben, wenn die L1 der Lernenden nicht eine zugrundeliegende SVO-Struktur besitzt (vgl. etwa die Studie von Haberzettl (2005), die im nächsten Kapitel genauer beschrieben wird).[59]

(Stufe 3) ADV: Focus position / Adverb Preposing

und am Sonntag *mein Bruder auch kommt nach Deutschland.*
(TN049, Zeitpunkt 1, Einladung nach Leipzig)

Auf Stufe 3 erwerben Lernende die Möglichkeit, das Adverb vor die kanonische SVO-Struktur zu positionieren. Dabei kann diese satzinitiale Fokusposition mit W-Fragewörtern, adverbialen Ergänzungen, Nominal- oder Präpositionalphrasen besetzt werden. Damit unterscheidet sich diese Stufe von der Stufe 2, da die leicht nachvollziehbare SVO-Struktur nicht mehr in diesem Maße gegeben ist. In dieser Form realisierte Sätze erfordern daher ein höheres Maß an Verarbeitungsressourcen (*processing ressources*, vgl. Pienemann 1998: 110). Obwohl es sich bei dieser Position um eine stark exponierte Position im Satz handelt, kann sie besetzt werden, ohne dass es in der folgenden SVO-Struktur zu einer Informationsanpassung, einer *feature unification*, kommen muss. Daher wird angenommen, dass diese Stufe vor den folgenden Stufen, die alle eine Informationsanpassung voraussetzen, erworben wird (vgl. Jansen 2008: 191).

59 An dieser Stelle ist darauf hinzuweisen, dass die PT zwar eine Phase der ‚kanonischen Wortstellung' annimmt, jedoch nicht genau spezifiziert, ob es sich bei dieser um eine SVO- oder eine (S)OV-Wortstellung handelt (vgl. Pienemann 1998: 313–315; vgl. Jansen 2008). Dies muss vielmehr in Einzelstudien jeweils aus den vorliegenden Daten herausgearbeitet werden. Insofern widerspricht eine von der SVO-Wortstellung abweichende Realisierung in dieser Phase aus Sicht der PT nicht den grundsätzlichen Annahmen, die in der *Processability Theory* getroffen werden (vgl. Pienemann / DiBiase / Kawaguchi 2005: 232).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

(Stufe 4) SEP oder PART: Split verbs / Verb separation

*Ich **lade** dich nach Leipzig **ein**.* (TN001, Zeitpunkt 1, Einladung nach Leipzig)

*Wir **können** auch das Völkerschlachtdenkmal **ansehen*** (TN004, Zeitpunkt 1, Einladung nach Leipzig)

Auf Stufe 4 müssen Lernende nun zum ersten Mal Informationen über Phrasengrenzen hinweg aneinander anpassen. Pienemann (2005a: 31) geht davon aus, dass auf dieser Stufe Verbalphrasen als eigenständige Konstituenten von den Lernenden wahrgenommen werden können. Gleichzeitig wird die initiale Hypothese von der Wortstellung von Verben, die bis dahin an der zweiten Stelle auftraten (außer bei einer initialen OV-Hypothese), überarbeitet, so dass Verben bzw. Verbteile oder Verbkomplementphrasen in einer finalen Position auftreten können. Im Deutschen betrifft dies Konstruktionen mit Hilfs- oder Modalverben (AUXILIARY) und die Vergangenheitskonstruktion mithilfe des Perfekts. Diese werden mit einem zusätzlichen lexikalischen Verb oder einem verbalen Partikel (PARTICIPLE) in Endstellung realisiert. Das Verb in Endstellung muss dabei nicht unbedingt korrekt realisiert werden (vgl. Pienemann 2005a: 31f.). In jedem Fall müssen Lernende zwischen diesen beiden Verbpositionen grammatikalische Informationen austauschen, was diese Stufe von der vorangegangenen Stufe strukturell unterscheidet und größere Verarbeitungsmöglichkeiten seitens der Lernenden voraussetzt (vgl. Jansen 2008: 192).

(Stufe 5) INV: Verb-second / Inversion

***endlich** essen wir zusammen sehr lecker syrisches Essen!* (TN029, Zeitpunkt 1, Einladung nach Leipzig)

Auf dieser fünften Stufe der PT-Hierarchie sind Lernende in der Lage, die satzinitiale Position mit etwas anderem als dem Subjekt zu besetzen und gleichzeitig die Verbzweitstellung beizubehalten. Damit unterscheidet sich diese Stufe etwa von Stufe 3, auf der einzelne Elemente lediglich vor eine darauffolgende SVO-Struktur gestellt werden konnten. Strukturell geht die PT hier davon aus, dass die Lernenden eine Regel erwerben, die ihnen die satzinitiale Position für eine optionale Verbposition links des Subjekts öffnet (vgl. Pienemann et al. 2005a: 33). Lernende, die wahrnehmen, dass sich nicht ein Subjekt, sondern ein anderes Element in der satzinitialen Fokusposition befindet, erkennen auf dieser Stufe, dass auf dieses Element eine obligatorische Inversion von Subjekt und finitem Verb folgen muss. Zum Teil ergaben Studien, die die Annahmen der PT auf schriftliche Lernerdaten adaptieren, für die Stufen 5 und 6 eine invertierte Reihenfolge, d.h. dass etwa Nebensätze mit Verbendstellung vor Sätzen mit Inversion erworben wurden (vgl. etwa Diehl et al. 2000). Die Ergebnisse dieser Studien müssen aus heutiger Sicht mit großer Vorsicht interpretiert werden (vgl. dazu das nächste Kapitel zu den methodischen Entwicklungen der PT).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

(Stufe 6) V-END: Verb-final in subordinate clauses

ich lade dich nach Leipzig ein, ***weil es so viel zu tun gibt!*** (TN029, Zeitpunkt 1, Einladung nach Leipzig)

Die höchste Stufe innerhalb der PT-Hierarchie bildet die Stufe der Verbendstellung in Nebensätzen. Bis zu diesem Zeitpunkt haben Lernende, laut Pienemann, die Wortstellung in Nebensätzen analog zur Wortstellung in Hauptsätzen realisiert (vgl. Pienemann 2005a: 34). Auf dieser Stufe können Lernende nun zwischen diesen beiden grundsätzlichen Worstellungsmustern differenzieren. Wesentliches Merkmal zum Nachweis dieser Stufe ist die Endstellung von Verben in Nebensätzen. Pienemann geht zusätzlich davon aus, dass auf dieser Stufe die Realisierung von Inversion in Nebensätzen graduell abnimmt (vgl. Pienemann 2005a: 34). Mehr noch, die Fähigkeit, Haupt- und Nebensätze strukturell voneinander zu unterscheiden, geht laut PT mit der Blockierung einer Inversion in Nebensatzkontexten einher. Um eine Inversion blockieren zu können, muss diese bereits erworben sein, weshalb die PT den Erwerb der Stufe V-End nach dem Erwerb der Stufe INV annimmt (vgl. Jansen 2008: 193).

Es wird grundsätzlich davon ausgegangen, dass Lernende diese Erwerbsstufen nacheinander und in der vorgegebenen Reihenfolge durchlaufen müssen, d.h., dass diese Stufen als implikationelle Sequenz angeordnet sind. Um Strukturen auf einer Stufe verarbeiten zu können, sind sowohl die Verarbeitungsprozeduren der jeweiligen Stufe als auch diejenigen aller darunterliegenden Stufen zwingende Voraussetzung (vgl. Pienemann 1998: 6). Die Emergenz von unterschiedlichen Stufen zur gleichen Zeit oder in einer anderen als der vorausgesagten Reihenfolge ist daher aus PT-Sicht nicht möglich (vgl. Jansen 2008: 206). Die PT macht keine Vorhersagen darüber, wann im zeitlichen Verlauf der zweitsprachlichen Entwicklung die beschriebenen Erwerbsstufen emergent werden und warum sie zu bestimmten Zeitpunkten emergieren (vgl. Jansen 2008: 193). Ebenfalls spielen die Frequenz der Anwendung von Strukturen oder deren Korrektheitsquote für die Voraussagen der PT keine Rolle (vgl. Jansen 2008: 204). Es wird angenommen, dass die Existenz dieser Stufen universell gilt, dass folglich unterschiedliche Ausgangssprachen (L1) oder unterschiedliche Erwerbskontexte (ungesteuert vs. Gesteuert; DaF vs. DaZ) oder der durch Unterricht vermittelte Input (*teachability hypothesis*, vgl. Pienemann 1998: 13) keinen Einfluss auf die grundsätzlich festgelegte Sequenz der Erwerbsstufen haben (vgl. Pienemann 1998; Jansen 2008: 194).[60] L1-L2-

60 Pienemann (1998) geht in seiner Ausformulierung der PT davon aus, dass die *teachability hypothesis* die Lehrbarkeit bestimmter Strukturen innerhalb der zweitsprachlichen Entwicklung stark einschränkt. Die Hypothese bezieht sich jeweils auf einzelne Lernende und ihre kognitive Bereitschaft (*readiness*), bestimmte Strukturen verarbeiten zu können. Pienemann nimmt an, dass durch gezielte Instruktion zwar die Reihenfolge des Erwerbs bestimmter Strukturen nicht verändert werden kann, dass aber sehr wohl die Variation innerhalb des Hypothesenraums und die Geschwindigkeit, mit der Strukturen erworben werden, beeinflusst werden können (vgl. u.a. Pienemann 2005a: 73).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Interferenzeffekte werden zwar nicht ausgeschlossen, die PT geht jedoch davon aus, dass erst dann Strukturen aus der L1 für mögliche Interferenzeffekte in der zweitsprachlichen Entwicklung relevant sein können, wenn Lernende strukturell ähnliche Elemente in der L2 verarbeiten können. Hier begrenzt die Verarbeitbarkeit ebenso mögliche Effekte (vgl. Pienemann / Lenzing 2015: 176). Selbst gezielter Input, der den vorausgesagten Stufen nicht entspricht, kann nicht dazu führen, dass Lernende etwa bestimmte Stufen überspringen oder hierarchisch höhere Stufen vor hierarchisch niedrigeren Stufen erwerben können. Als Beleg für die Emergenz einer Struktur, als Emergenzkriterium, können bereits wenige Belege, die sich auf die oben beschriebenen Kernindikatoren (*key grammatical encoding phenomenon*; Pienemann 1998: 13) innerhalb der lernersprachlichen Strukturen beziehen, herangezogen werden. Pienemann (1998: 145) setzt als Emergenzkriterium die einmalige (zielsprachliche) Realisierung einer Struktur in mindestens vier für diese Struktur obligatorischen Kontexten an. Auf ähnliche Weise sieht er eine Stufe als nachweislich nicht-erworben an, wenn sich vier für eine Struktur obligatorische Kontexte finden, ohne dass die Struktur ein einziges Mal korrekt verwendet wurde (vgl. Jansen 2008: 202). Emergenz wird selbst in PT-basierter Forschung nicht durchgehend mit demselben Emergenzkriterium operationalisiert (vgl. Lenzing et al. 2019a: 4), was durchaus dazu führen kann, dass der Zeitpunkt der Emergenz einer Struktur unterschiedlich bestimmt werden kann (vgl. hierzu den Abschnitt zu Kritik und Grenzen in Kapitel 2.4.4). Lenzing (2015) wendet etwa ein Emergenzkriterium von drei Vorkommen einer Struktur pro Lernertext an. Ihr Vorgehen ähnelt damit dem Kriterium, das in der Profilanalyse nach Grießhaber (vgl. Kapitel 2.4.3) operationalisiert wird.

Darüber hinaus wird innerhalb der PT davon ausgegangen, dass es auf jeder Entwicklungsstufe innerhalb dieser implikationellen Sequenz einen großen Spielraum („leeway", Pienemann / Lenzing 2015: 164) für Variation innerhalb der Lernersprache der Lernenden geben kann. Dieser Spielraum konstituiert einen Hypothesenraum (*hypothesis space*), innerhalb dessen eine bestimmte Menge an Lernergrammatiken realisiert werden kann (Pienemann 1998: 10; vgl. zur *Interlanguage Variation* nach Pienemann zusätzlich Meisel / Clahsen / Pienemann 1981). Beobachtbare interindividuelle Variation kann folglich als der individuelle Gebrauch von begrenzten strukturellen Optionen beschrieben werden (Pienemann 1998: 10). Dieser Raum wird als Zylinder visualisiert, wobei die zweitsprachliche Entwicklung anhand unterschiedlicher linearer Entwicklungen innerhalb dieses Zylinders nachvollzogen werden kann (vgl. Pienemann 1998: 232). Die vertikale Ebene innerhalb dieses Zylinders verdeutlicht die beschriebenen Entwicklungsstufen und die horizontale die jeweils mögliche Variation auf dieser Stufe. Mit fortschreitendem Erwerb der einzelnen Stufen wird folglich der Raum größer, in dem Lernersprache variieren kann (vgl. Lenzing 2015: 95). Dies hat mit jeder erworbenen Stufe einen potenziell größeren Raum zur Realisierung variabler Grammatiken zur Folge (vgl. Lenzing 2015: 95), wobei individuelle lernersprachliche Variation innerhalb dieses Raums nicht vorausgesagt werden kann. Gleichzeitig können auf jeder Entwicklungsstufe in

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

einem insgesamt begrenzten Raum, eben dem Hypothesenraum, nur eine endliche Anzahl an Möglichkeiten realisiert werden. PT-basierte Ansätze nehmen so gleichzeitig eine regelbasierte zweitsprachliche Entwicklung an und versuchen der für Lernersprache charakteristischen Variation gerecht zu werden (vgl. Pienemann / Lenzing 2015: 160). Gerade durch die Arbeiten von Dyson (2021; 2022) wird die Diskussion um Variation innerhalb des PT-Kontextes zudem in jüngster Zeit wieder verstärkt ins Zentrum der Diskussion gerückt.

Methodisch existieren innerhalb der PT vor allem zwei Methoden, die zur Überprüfung der Hypothesen und Voraussagen der PT angewendet werden. Einerseits kommt das von Clahsen (1985) entwickelte und innerhalb des PT-Kontextes weiterentwickelte *Rapid Profiling* (vgl. Keßler / Liebner 2011) zum Einsatz, das über den Einsatz einer spezialisierten Software die schnelle Einstufung lernersprachlicher Äußerungen erlaubt, aber bis heute m. W. nur für die Analyse von Englisch als L2 angewandt und nicht für das Deutsche als L2 adaptiert wurde. Andererseits setzen viele Studien Implikationsskalen ein, um die implikationell-sequenziellen Aspekte des L2-Erwerbs herauszuarbeiten (vgl. für eine genaue Darstellung der Funktionsweise von Implikationsskalen Fußnote 56 und die Erklärung bei Jansen 2008). Im deutschsprachigen DaF/DaZ-Kontext werden oft Korrektheitswerte analysiert, ohne diese auf einer Implikationsskala anzuordnen (vgl. dafür den folgenden Abschnitt zur methodischen Entwicklung von PT-basierten Studien, Kapitel 2.4.2).

2.4.2 Empirisch-methodische Entwicklung der *Processability Theory*

Im Folgenden werden Studien aus dem PT-Forschungskontext vorgestellt, die von besonderer Relevanz für die vorliegende Arbeit sind. Die ausgewählten Studien beziehen sich alle auf das Deutsche als Fremd- oder Zweitsprache.[61] Alle diese Studien untersuchen den Erwerb bzw. die Entwicklung der deutschen Wortstellung.[62] Außerdem wurden schwerpunktmäßig Studien herangezogen, die sich auf schriftliche Lernersprachdaten stützen. Studien, die mit mündlichen Daten arbeiten, wurden nur einbezogen, falls diese aus Sicht des Verfassers einen besonderen Beitrag zur methodischen Weiterentwicklung oder zur Überprüfung von Hypothesen der PT in einem anderen als dem ursprünglichen Erwerbskontext (ungesteuerter DaZ-Erwerb erwachsener Lernender) geliefert haben. Aus dieser Auswahl ergibt sich ein begrenzter, aber für die vorliegende Studie relevanter Einblick in ausgewählte Teilbereiche.

61 Das bedeutet im Kontext der PT eine relativ starke Einschränkung, da PT-basierte Studien neben dem Deutschen vor allem das Englische als L2 untersucht haben. Gleichzeitig würde ein Einbezug dieser Studien den Umfang der vorliegenden Studie deutlich erweitern und den Fokus unter Umständen verschieben.

62 In vielen Studien wurde neben der Wortstellung des Deutschen unter anderem der Kasuserwerb aus einer PT-Perspektive untersucht (beispielhaft hier etwa die Arbeiten von Kristof Baten, vgl. Baten 2010, 2011, 2013 und 2014).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

In einer frühen Studie zum sequenziellen Erwerb der deutschen Wortstellung in einem gesteuerten Erwerbskontext kann Ellis (1989) mithilfe der Daten von 39 (ebd.: 312) erwachsenen und gerade beginnenden Lernenden des Deutschen als Fremdsprache mit unterschiedlichen L1 ähnliche Erwerbssequenzen wie im ungesteuerten Zweitspracherwerb zeigen (ebd.: 319). Die erreichten Erwerbsstufen werden schneller und erfolgreicher von den Lernenden im gesteuerten Erwerb durchlaufen. In seiner Studie überprüft Ellis mithilfe einer zweimal in Partnerarbeit durchgeführten Bildbeschreibungsaufgabe (ebd.: 315) die Hypothese, dass Instruktion die Erwerbssequenz syntaktischer Strukturen nicht entscheidend beeinflussen kann (ebd.: 307). Hier arbeitet Ellis Erwerbssequenzen heraus, die von den im Unterricht präsentierten Sequenzen abweichen (ebd.: 319f.). Zur Überprüfung der Ergebnisse führt Ellis zwei Implikationsanalysen durch, eine Querschnittsanalyse, die sich nur auf den zweiten Zeitpunkt bezieht, und eine ‚longitudinale' Analyse, die die Lernenden einbezieht, die zu beiden Zeitpunkten ausreichend obligatorische Kontexte für bestimmte Erwerbsstufen produzieren (ebd.: 317). Ellis setzt ein Erwerbskriterium von mindestens drei korrekt realisierten Strukturen in mindestens vier obligatorischen Kontexten an, oder anders ausgedrückt eine Korrektheitsquote von 75 %.[63]

Eine für den PT-Framework klassische Studie ist die Arbeit von Pienemann (1998), die gleichzeitig in vielen Bereichen die theoretischen Grundlagen der PT liefert und bis heute die mit Abstand umfangreichste Darstellung des Erklärungsansatzes darstellt. Pienemann untersucht in einer longitudinalen Einzelfallstudie den Spracherwerb des australischen Studenten und DaF-Lernenden Guy, der ohne Vorkenntnisse begonnen hatte, Deutsch zu lernen (vgl. Pienemann 1998: 117). Zur Datenerhebung werden über die Dauer von einem Jahr alle vierzehn Tage mündliche Interviews durchgeführt. In der Studie untersucht Pienemann die Bereiche Wortstellung und Morphologie. Pienemann kann für Guy, der in einem gesteuerten Kontext Deutsch als Fremdsprache lernt, herausarbeiten, dass der Erwerb der Strukturen trotz abweichenden Inputs sequenziell verläuft und dabei derselben Sequenz folgt, die innerhalb der ZISA-Studie für ungesteuerte Erwerbskontexte herausgearbeitet wurde (ebd.: 122).

63 Zusätzlich untersucht Ellis Gruppenunterschiede in Bezug auf die fünf beobachteten Lerngruppen in seiner Studie. In diesen Gruppen wurden die verschiedenen syntaktischen Strukturen (PARTICLE, INVERSION, VERBEND) unterschiedlich intensiv behandelt. Trotz allem ergeben sich die jeweils höchsten Korrektheitsquoten für PARTICLE und die geringsten für VERBEND. Obwohl in zwei Gruppen ein Schwerpunkt auf die Vermittlung von INVERSION gelegt wurde, ist die Korrektheitsquote für PARTICLE in diesen Gruppen deutlich höher (ebd.: 321). Ellis berichtet für eine Gruppe den Erwerb von VERBEND-Strukturen, während sich gleichzeitig kaum Belege für INVERSION finden. Leider kann sich Ellis diesen Faktor nicht erklären. Diese Beobachtung erscheint rätselhaft, da es sich bei dieser Gruppe um die Gruppe mit der geringsten Sprachlernmotivation zu handeln scheint.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Eine ebenfalls für den deutschen Forschungskontext relativ frühe, wenngleich bis heute oft zitierte Studie stellt die querschnittlich (bzw. pseudolongitudinal) angelegte DiGS-Studie (Deutsch in Genfer Schulen) von Diehl et al. (2000) dar (vgl. Diehl 1998; Diehl et al. 2002).[64] Die DiGS-Studie besitzt im Vergleich zu vielen anderen Studien eine relativ breite Datenbasis. Insgesamt werden in einem klassischen DaF-Kontext 1800 Texte von 300 Schülerinnen und Schülern (SuS) aus unterschiedlichen aufeinander folgenden Klassenstufen an Genfer Gymnasien untersucht.[65] Damit gehört das Korpus der DiGS-Studie bis heute zu den umfangreichsten deutschsprachigen Lernerkorpora. Diehl et al. (2000) rekonstruieren den Erwerb der deutschen Satzmodelle und der Morphologie des Deutschen. Die Autorinnen und Autoren arbeiten in ihrer Studie zum gesteuerten Erwerb des Deutschen in Bezug auf die Wortstellung Sequenzen heraus, die leicht von den Sequenzen der PT abweichen. VEND wird von den Genfer SuS vor INV erworben. Methodisch werden nicht einzelne SuS in ihrer Entwicklung betrachtet, sondern nur die Durchschnittswerte verschiedener Klassenverbände berechnet und daraus universelle Erwerbsstufen abgeleitet. Dies ist aus methodischer Perspektive durchaus problematisch und zusätzlich nicht mit dem emergenzbasierten Ansatz der PT vereinbar, der grundsätzlich von einzelnen Lernenden bzw. von Einzeltexten ausgeht, um die Emergenz von Strukturen zu belegen.

Eine Studie aus dem schwedischen DaF-Kontext legt Håkannson (2001) vor.[66] Sie untersucht auf Grundlage der durch die PT vorausgesagten Sequenzen die

64 Studien werden als pseudolongitudinal beschrieben, wenn Daten von Lernenden lediglich zu einem Zeitpunkt gesammelt werden und diese Daten dann in Beziehung gesetzt werden zu vergleichbaren Daten von Lernenden auf höheren Niveaustufen oder wie in diesem Fall in höheren Klassenstufen. Ziel eines solchen Studiendesigns ist es, Entwicklungen abzubilden, die in einem rein querschnittlichen Design nicht gezeigt werden könnten.

65 Die Texte der DiGS-Studie sind im PDF frei verfügbar und unter https://www.unige.ch/lettres/alman/de/recherche/abgeschlossene-projekte/digs/ abrufbar [23.02.2023].

66 Die Studie von Håkannson et al. (2002) wird an dieser Stelle nicht in die Diskussion miteinbezogen, da es sich einerseits um die Elizitation mündlicher Sprachdaten handelt und andererseits eines der wichtigen Ziele der Studie die Untersuchung möglicher Transfereffekte zwischen L1 und L2 ist. Damit weist die Studie einen theoretischen Fokus auf, der für die vorliegende Studie nicht unmittelbar von Relevanz ist. Ähnliches gilt für die Studie von Bohnacker (2006), die mündliche Sprachdaten von sechs erwachsenen Lernenden mit Schwedisch als L1 erhebt, wobei drei Lernende Englisch als Fremdsprache gelernt haben und drei Deutsch als erste Fremdsprache lernen. Bohnacker untersucht einerseits den Erwerb der Verbzeitstellung, wobei sie konstatiert, dass die von ihr beobachteten Lernenden mit dieser keine Schwierigkeiten haben (ebd.: 477) und andererseits mögliche Transfereffekte aus den bereits beherrschten Sprachen, die dem Erwerb des Deutschen vorausgehen. Hier geht Bohnacker davon aus, dass ihre Daten Transfereffekte aus der L1 stützen und Kenntnisse in einer weiteren Fremdsprache bestimmte Effekte unter Umständen verschleiern können (ebd.: 478).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Hypothese bestimmter Transfereffekte unterschiedlicher Erstsprachen beim Erwerb des Deutschen als L2. Die 143 Teilnehmenden (vgl. Håkannson 2001: 8) ihrer Querschnittstudie sprechen Schwedisch als L1 und lernen Deutsch als weitere Fremdsprache nach Englisch (vgl. Håkannson 2001: 6). Alle Teilnehmenden sind SuS an schwedischen Schulen (vgl. ebd.: 8) zwischen der 7. und der 9. Klasse, d.h. sie sind 13–15 Jahre alt. Die Lernenden sollen einerseits Sätze übersetzen und andererseits Lücken bei bereits konstruierten Sätzen ausfüllen (vgl. ebd.: 9). Håkannson bestätigt in ihren Ergebnissen die durch die PT vorausgesagten Sequenzen (ebd.: 14). Die von Håkannson herausgearbeiteten Ergebnisse deuten darauf hin, dass die beobachteten Strukturen nicht aufgrund von Transfereffekten aus dem Schwedischen oder dem Englischen zustande kommen, sondern dass ihre Ergebnisse vielmehr die Annahmen der PT zu stützen scheinen (vgl. ebd.: 14). Sie macht jedoch an mehreren Stellen auf die sehr große interindividuelle Variabilität zwischen den Lernenden und zusätzlich auf große intraindividuelle Unterschiede aufmerksam (vgl. etwa ebd.: 9 oder 13). Da alle Teilnehmenden exakt dieselben Sätze bearbeiten, kann Håkannson diesen Faktor gut beobachten.[67] Zu beachten ist ebenfalls, dass es sich nicht um gesprochen- oder spontansprachliche Daten handelt, sondern um schriftliche Daten, die in einem stark eingegrenzten Aufgabenaufbau erhoben werden. Die Daten aus der Studie von Håkannson sind daher nur bedingt mit anderen hier beschriebenen Studien vergleichbar und haben aufgrund ihres Querschnittcharakters nur eine sehr begrenzte Aussagekraft, was die Entwicklung implikationeller Sequenzen angeht.

Einen zum Teil PT-kritischen Beitrag aus der DaF-Perspektive stellt die Studie von Meerholz-Härle / Tschirner (2001) dar. Die beiden untersuchen in einer Querschnittsstudie sechs ACTFL *Oral Proficiency Interviews* (OPIs), die mit Studierenden mit Englisch als L1 aus den USA durchgeführt und auf dem Niveau *Intermediate Mid* bewertet wurden (ebd.: 163f.). Dabei wollen sie die Frage beantworten, ob „die Subjekt-Verb-Inversion auf unterschiedlichen Erwerbsstufen erworben wird, je nachdem, ob sie in Frage- oder in Aussagesätzen auftritt" (ebd.). In den Transkriptionen bestimmen die Autor/-innen die obligatorischen Inversionskontexte, die sie nach „Wortfragen, Satzfragen und Aussage" unterscheiden (ebd.: 164). Aus den zielsprachlich und den nicht-zielsprachlich realisierten Kontexten wird ein Korrektheitsquotient berechnet. Anhand dieser Korrektheitswerte leiten Meerholz-Härle / Tschirner ab, dass die Inversion in Aussagesätzen für die untersuchten Lernenden herausfordernder ist als in Fragesätzen, die durchgehend zielsprachlich korrekt realisiert werden (ebd.: 166). Die beiden weisen als methodische Herausforderung darauf hin, dass diese Unterschiede nicht mehr sichtbar sind, sobald mit den Gesamtwerten der Inver-

67 Leider finden sich in Håkannsons Bericht nur wenige deskriptive Statistiken oder die Rohdaten der einzelnen Teilnehmenden, die eine genauere Einschätzung und Analyse der Variabilität in den Daten zulassen würden.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

sionskontexte gearbeitet wird und dass daher sowohl ein umsichtiges Vorgehen auf der Ebene der einzelnen Erwerbsstufen als auch bei der Verwendung von Mittelwerten angebracht ist (vgl. ebd.: 167). Die Ergebnisse von Meerholz-Härle / Tschirner stellen aus Sicht der beiden die implikationell-sequenzielle (bei Meerholz-Härle / Tschirner: serielle) Abfolge der Erwerbsstufen in Frage. Vielmehr scheinen in ihren Daten bestimmte Entwicklungen parallel abzulaufen. Tatsächlich können aufgrund des querschnittlichen Designs hier keine (oder nur sehr begrenzte) Rückschlüsse auf zweitsprachliche Entwicklungsprozesse gezogen werden.

Ebenfalls aus einer DaF-Perspektive untersucht Boss (2004) insgesamt „60 [schriftliche, MS] Texte von 60 Testpersonen" (ebd.: 29), die an drei verschiedenen australischen Universitäten einen Deutschkurs besuchen und Englisch als L1 sprechen. Für die longitudinale Studie werden von den Testpersonen insgesamt fünf verschiedene Texte verfasst, jedoch nicht an jeder Hochschule alle fünf. Die syntaktische Progression im Unterricht ist an allen Hochschulen dieselbe (SVO + Fragesätze → Inversion → Separation → Verbendstellung) (vgl. ebd.). Boss interessiert sich vor allem für die beiden Erwerbsstufen Inversion und Verbendstellung. In den Texten der Lernenden identifiziert sie satzwertige Einheiten und annotiert zielsprachlich realisierte Strukturen und obligatorische Kontexte der beiden Stufen Inversion und Verbendstellung. Darauf aufbauend berechnet sie Korrektheitsquoten, die sie aggregiert für die einzelnen Gruppen an den drei Universitäten und aufgeteilt nach Texten angibt (ebd.: 30).[68] Ein Emergenzkriterium legt sie, da Gruppendaten betrachtet werden, konsequenterweise nicht an. Boss kommt anhand der von ihr berechneten Korrektheitsquoten zu dem Ergebnis, dass Inversion und Nebensätze zwar nicht gleichzeitig emergieren, was sie mit der Progression im Unterricht erklärt, aber zeitgleich ‚erworben' werden (vgl. ebd.: 32).[69] Hier übernimmt sie die Terminologie aus Diehl et al. (2000). Kritisch anzumerken ist, dass Boss keinerlei deskriptive Statistik oder Hinweise zur Verteilung der Daten an den einzelnen Universitäten zur Verfügung stellt. Darauf, dass ihr Befund des zeitgleichen Erwerbs in Kontrast zu den Voraussagen der PT steht, geht Boss nicht weiter ein.

68 Im Gegensatz zu Meerholz-Härle / Tschirner (2001) schließt Boss (2004) Fragesätze mit Inversion aus ihrer Analyse aus und betrachtet nur Inversionen in Aussagesätzen.

69 Zu einem ähnlichen Ergebnis wie Boss (2004) gelangt Lund (2004), der einerseits mündliche Daten von 21 DaF-Studierenden mit L1 Englisch aus dem amerikanischen Kontext (vgl. Lund 2004: 100) auswertet und andererseits metasprachliches Wissen der Studierenden testet, indem diese nicht-zielsprachliche Sätze korrigieren müssen. Seine Ergebnisse lassen ihn zu dem Schluss kommen, dass die beiden Stufen Inversion und Verbendstellung nicht in einem sequenziellen Verhältnis zueinanderstehen und „entweder gleichzeitig oder in individuell variierenden Folgen erworben bzw. beherrscht werden können" (ebd.: 102).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Mit einer kritischen Perspektive auf verschiedene Erwerbstheorien (nämlich aus Sicht der UG und PT) betrachtet Haberzettl (2005) im Rahmen von vier longitudinalen DaZ-Einzelfallstudien den weitgehend ungesteuerten Erwerb der Verbstellungsregeln im Deutschen als L2 bei Kindern mit Türkisch und Russisch als L1 im Alter von sechs bis acht Jahren (vgl. ebd.: 77).[70] Die mündlichen Daten werden in einem Zeitraum von 2–4 Jahren in monatlich durchgeführten, freien Gesprächen und durch die Nacherzählung von Bildergeschichten erhoben. Haberzettl geht den Fragen nach, wie die vier Kinder die Verbstellungsregeln erwerben und ob sich ein Einfluss der jeweiligen L1 in den Erwerbssequenzen feststellen lässt (vgl. ebd.: 2). Sie arbeitet Unterschiede in Bezug auf die zweitsprachliche Entwicklung der Verbstellungsregeln zwischen den Kindern mit Türkisch als L1 und den Kindern mit Russisch als L1 heraus, die sie unter anderem am unterschiedlich schnellen Erwerb der deutschen Verbstellungsregeln festmacht. Während zu Beginn der Erhebung von den russischsprachigen Kindern bereits korrekte Verbzweitstellungen produziert werden, finden sich bei den türkischsprachigen Kindern zahlreiche nicht-zielsprachliche Formulierungen (ebd.: 135f.). Im Folgenden fällt es den türkischsprachigen Kindern jedoch leichter, die übrigen Satzstellungsmuster (SEP, INV, VEND) zu entwickeln. Haberzettl geht davon aus, dass sie schneller die Hypothese von Deutsch als reiner SVO-Sprache verwerfen können, weil das Türkische als L1 eine SOV-Sprache darstellt, während die Kinder mit Russisch als L1 hier größere Schwierigkeiten haben (vgl. ebd.).[71] Hinsichtlich der Anwendbarkeit ihrer Ergebnisse im Rahmen der PT geht Haberzettl davon aus, dass sich die russischen Lernenden in ihren Erwerbsverläufen mit den von Pienemann (1998) herausgearbeiteten Sequenzen decken, sich für die türkischen Lernenden diese Sequenzen jedoch nicht bestätigen lassen. Diese gehen initial nicht von einer SVO-Struktur aus und beginnen den Erwerb des Deutschen als Zweitsprache mit einer OV bzw. (S) OV-Hypothese (vgl. Haberzettl 2005: 159f.). Darüber hinaus geht sie davon aus, dass die von ihr beobachteten russischen Lernenden die Stufe INV (Inversion) vor der Stufe (SEP) Separation erwerben (vgl. ebd.).

Ballestracci (2008) untersucht in ihrer pseudolongitudinalen Studie anhand von 157 Texten zu acht Zeitpunkten den Erwerb der Verbalmorphologie, der Satzgliedstellung und der Deklination von Nominalphrasen bei italienischen DaF-Studierenden, die zuvor kein Deutsch lernten. Wie Diehl et al. (2000) versucht Ballestracci überindividuelle Erwerbsphasen beim Erwerb des Deutschen als L2

70 Haberzettl benutzt für diese Studie das Augsburger Korpus (Wegener 1988–1992), das zwischen 1989 und 1992 von Heide Wegener erhoben wurde und in dem sich longitudinale Daten von insgesamt 12 Kindern mit den Erstsprachen Russisch, Türkisch und Polnisch finden. Weitere Informationen und eine für den Kontext DaF/DaZ ebenfalls relevante Untersuchung zum Erwerb der Nominalflexion finden sich in Wegener (1995).

71 Tatsächlich sagt die PT für den Fall, dass die L1 eine SOV-Sprache darstellt, einen unterschiedlichen Erwerbsverlauf voraus (vgl. Pienemann 1998, 2005).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

herauszuarbeiten. Da nicht für alle Studierenden longitudinale Daten vorliegen, werden die Daten aller Studierenden in chronologische Phasen eingeteilt und aggregiert analysiert (vgl. Ballestracci 2008: 163–164). Dies bedeutet, dass, wie bei der Studie von Diehl et al. (2000), nicht mehr die einzelnen Lernverläufe in den Blick genommen werden. Die Analyse beruht auf dem Vergleich obligatorischer Kontexte und der in diesen ‚normkonform' (vgl. Ballestracci 2008: 164) produzierten Strukturen. Ein Emergenzkriterium wird nicht in die Analyse einbezogen und die Daten stehen nicht öffentlich zur Verfügung. Ballestracci weist darauf hin, dass in ihren Daten „nicht immer von klaren Erwerbsfolgen auszugehen ist, da die italophonen Lerner auch individuelle Variation aufweisen, die bei der Generalisierung der Daten verloren geht" (Ballestracci 2008: 162). Sie arbeitet zudem Erwerbsphasen heraus, innerhalb derer aus ihrer Sicht grammatische Strukturen gleichzeitig erworben werden (vgl. ebd.: 164). Dies steht allerdings in Konflikt zu den Annahmen der PT, was jedoch nicht weiter reflektiert wird.

Eine weitere Querschnittsstudie aus dem DaF-Kontext ist die Studie von Jansen (2008). Sie untersucht die Daten von insgesamt 21 australischen Studierenden mit Englisch als L1, die in 45-minütigen und nicht weiter angeleiteten Interviews mit L1-Sprechenden des Deutschen aufgezeichnet wurden.[72] Die Autorin möchte unter anderem die Hypothese überprüfen, ob sich in einem gesteuerten bzw. instruierten DaF-Lernkontext dieselben Erwerbssequenzen wie bei den Lernenden aus der angesprochenen ZISA-Studie zeigen lassen (vgl. ebd.: 194). Dabei werden in den Transkripten minimale satzwertige Einheiten (*clauses*; ebd.: 196) annotiert und diese nach der Stellung von Subjekten, Verben und (adverbialen) Ergänzungen klassifiziert. Die einzelnen Lernenden werden anhand von Implikationsskalen weiter untersucht, um festzustellen, welche Strukturen aus den Daten zu rekonstruieren waren (ebd.: 198). Jansen wendet Pienemanns Emergenzkriterium von einer korrekten Realisierung einer Struktur in mindestens vier obligatorischen Kontexten an (vgl. ebd. 202). Sie sammelt nicht nur Belege für die Emergenz von Strukturen, sondern zusätzlich Belege dafür, dass bestimmte Strukturen noch nicht emergiert sind, womit sich ihre Studie von vielen der bis zu diesem Punkt vorgestellten Studien unterscheidet. Dies ist aus ihrer Sicht dann der Fall, wenn sich, analog zum Kriterium für Emergenz, keine korrekte Realisierung einer Struktur in mindestens vier obligatorischen Kontexten finden. Jansen geht davon aus, dass vor allem der Nachweis von noch nicht erworbenen Strukturen bei gleichzeitigen Belegen für die Emergenz der direkt darunterliegenden Struktur die PT-Hypothese von der sequenziellen Abfolge der Erwerbsstufen stützt. In den Daten müsste sich beispielsweise die Emergenz von SEP nachweisen lassen, während sich gleichzeitig Belege für die Nichtemergenz von INV finden lassen. Die herausgearbeiteten Implikationsskalen unterstützen die Annahmen über den implikationellen Verlauf anhand der von der PT vorausgesagten Erwerbsstufen laut Jansen in eindrücklicher Weise.

72 Die Vorgabe der Interviews war „to get to know one another" (Jansen 2008: 195).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Sie spricht von einem „perfect fit" (ebd.: 207). Für alle Lernenden zeigen sich die vorausgesagten Sequenzen bzw. Belege dafür, dass sich die Lernenden auf bestimmten Erwerbsstufen (und nur auf diesen) befinden (ebd.: 205). Jansen macht darauf aufmerksam, dass sie nur für sieben von 21 Lernenden Belege dafür finden kann, dass bestimmte Stufen im Vergleich zu benachbarten Stufen noch nicht erworben sind (vgl. ebd.: 207). Dies schränkt aus ihrer Sicht und in ihrem Datenset die Belege für einen sequenziellen Verlauf etwas ein (ebd.: 207f.). Sie macht darauf aufmerksam, dass es sich hier um fehlende Belege für die Nicht-Emergenz einer Struktur und nicht etwa um einen Gegenbeleg der von der PT aufgestellten Hypothese zur sequenziellen Abfolge der Erwerbsstufen handelt. Einschränkend muss zusätzlich gesagt werden, dass Jansen eine Querschnittsstudie durchführt, die nicht die Entwicklungsverläufe von einzelnen Lernenden nachzeichnet, sondern, wie in der ZISA-Studie, für eine Gruppe eine bis dahin homogene und vergleichbare Entwicklung annimmt, was die Verwendung einer Implikationsskala auf Gruppenebene rechtfertigt. Um hier präzisere Aussagen treffen zu können, wären longitudinale Daten der einzelnen Lernenden notwendig.

Czinglar (2014) untersucht in einer longitudinalen Studie über 18 Monate den ungesteuerten Zweitspracherwerb zweier Schwestern mit der L1 Russisch und der L2 Deutsch. Der Fokus der Untersuchung liegt auf dem Alter als Faktor in der zweitsprachlichen Entwicklung hinsichtlich der Entwicklung von Verbstellungsmustern des Deutschen. Die beiden Schwestern sind zu Beginn der Studie 8 (Nastja) und 14 (Dascha) Jahre alt. Czinglar greift auf einstündige Sprachsamples spontaner und gesprochener Sprache zurück, die einmal pro Woche während der Interaktion mit L1-Deutsch-Sprechenden Interviewpartnerinnen und Interviewpartnern aufgezeichnet werden. Die Daten entstammen dem DaZ-AF-Korpus (Deutsch als Zweitsprache–der Altersfaktor) (vgl. Czinglar 2014: 106).[73] Insgesamt wertet Czinglar 42 Aufnahmen aus (21 von Nastja und 21 von Dascha).[74] Die Daten sind sehr gut vergleichbar, da es sich um Geschwister aus demselben familiären Umfeld handelt, die im selben Haushalt leben und denselben Bildungshintergrund besitzen. Beide Schwestern besuchen in Deutschland direkt den Regelunterricht und nicht etwa einen vorbereitenden DaZ-Unterricht, weshalb ihr Spracherwerb von Czinglar als eher ungesteuert eingestuft wird (vgl. ebd.: 106). Unterschiede ergeben sich vor allem im präbzw. postpubertären Alter der beiden Lernerinnen (vgl. ebd.: 109). Für ihre Analysen wählt die Autorin drei „zentrale Problembereiche der Verbstellung […]: die OV/XV-Stellung in der Satzklammer, die V2-Stellung mit Inversion und die VE-Stellung im Nebensatz" (ebd.: 248). Für die sehr umfangreichen Daten definiert Czinglar eine Korrektheitsquote von 90 % in allen obligatorischen Kontex-

73 Das Korpus ist auf den Seiten des *Language Archive* des Max-Planck-Instituts für Psycholinguistik in Nijmegen verfügbar.

74 Im Korpus stehen 128 Aufnahmen zur Verfügung, 65 von Dascha und 63 von Nastja (vgl. Czinglar 2014: 106).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

ten als Erwerbskriterium und eine Korrektheitsquote von 50 % als Kriterium für den beginnenden Erwerb einer zielsprachlichen Struktur (vgl. ebd.). Czinglar kann herausarbeiten, dass die Erwerbsgeschwindigkeit mit zunehmendem Alter tendenziell abnimmt (vgl. ebd. 179). Beide Lernerinnen stellen sich in Czinglars Analysen als sehr schnelle und erfolgreiche Lernerinnen heraus, wobei die jüngere Lernerin am Ende der Untersuchungszeit „den Verbstellungserwerb bereits vollständig abgeschlossen" (ebd.: 247) hat. Sie beginnt darüber hinaus „mit dem Erwerb aller drei Strukturformate ungefähr gleichzeitig, was die Vermutung nahelegt, dass sie einen systematischen Zusammenhang zwischen den drei Lerngegenständen herstellt" (ebd.: 248). Für die ältere Lernerin gilt dies nicht. Am Ende des Erhebungszeitraums ist nicht ganz klar, ob alle Verbstellungsregeln erworben wurden (vgl. ebd.). Ihr Verbstellungserwerb ähnelt, laut Czinglar, eher dem Erwerbsverlauf erwachsener Lernender, selbst wenn dieser deutlich schneller abläuft als bei vergleichbaren Erwachsenen (vgl. ebd.). Zentrale Desiderata aus Czinglars multifaktorieller Analyse beziehen sich auf den systematischen Einbezug weiterer „außersprachliche[r] Einflussfaktoren wie Motivation, Sprachtalent und Menge und Qualität des Inputs [neben dem Alter bei Erwerbsbeginn; MS]" (ebd. 252). Ebenso sollte aus ihrer Sicht die L1 stärker in Erwerbsstudien miteinbezogen werden. Ein wichtiges Forschungsfeld sieht Czinglar darüber hinaus im Bereich der Erwerbsgeschwindigkeit (vgl. ebd.).

Winkler (2017) untersucht in ihrer Dissertation mithilfe einer Interventionsstudie in einem DaF-Kontext erwachsene und gerade beginnende Deutschlernende mit Italienisch als L1 im Rahmen eines Kurses mit insgesamt 60 Unterrichtseinheiten. Sie erhebt die Daten von insgesamt 43 Studierenden (vgl. Winkler 2017: 153). Diese Studierenden werden in zwei Gruppen aufgeteilt, wobei eine Gruppe (n=21) nach ‚traditionellem' Curriculum und eine Gruppe (n=22) nach einem von Winkler kleinteilig entwickelten ‚naturalistischen' (bzw. natürlichen) Curriculum Deutsch lernen. Die Beschreibung ‚natürlich' bezieht sich in Winklers Studie vor allem auf die angenommene natürliche Erwerbsreihenfolge in der Sprachentwicklung der L1. Der Testgruppe mit dem ‚natürlichen' Curriculum werden daher vor allem OV-Strukturen vermittelt, während in der Kontrollgruppe mit ‚traditionellem' Curriculum der Schwerpunkt auf Deutsch als Verbzweit-Sprache (SVO) liegt (vgl. die tabellarische Übersicht der beiden Curricula in Winkler 2017: 354–366). Winkler geht davon aus, dass kontrollierter Input in Form von speziell aufbereiteten Lese- und Hörtexten sowie Übungsaufgaben mehr Einfluss auf die Erwerbsreihenfolge syntaktischer Strukturen als explizite Erklärungen bestimmter Strukturregeln haben könnte (vgl. ebd.: 161f.). Ihre Hypothese überprüft Winkler mithilfe von mehrfach wiederholten Tests, aber nicht anhand von spontan oder frei produzierten lernersprachlichen Daten. Die durchgeführten Tests sind viermal wiederholte, schriftliche Satzpuzzle-Tests, bei denen die Lernenden korrekte Sätze aus einer Auswahl von verschiedenen Wörtern rekonstruieren müssen (ebd.: 177) und Tests, bei denen mündlich Aktivitäten benannt werden sollen, die den Lernenden auf Bildern präsentiert werden (vgl. ebd.: 183). Außerdem werden *Elicited Imitation*-Tests durchgeführt,

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

bei denen die Lernenden einen bestimmten (gehörten) Satz mündlich wiedergeben sollen (ebd.: 185). Der Benennungstest und der *Elicited Imitation*-Test werden jeweils zweimal durchgeführt. Winkler kann anhand ihrer Daten zeigen, dass die Intervention erfolgreich war, da die Testgruppe, der eher die OV-Struktur des Deutschen vermittelt wird, nach ~60 UE weniger Schwierigkeiten mit der Produktion der deutschen Satzklammer hat, während die Satzklammer die Lernenden in der Kontrollgruppe noch vor große Herausforderungen stellt (vgl. ebd.: 316f.). Gleichzeitig hat die unterschiedliche Instruktion keinen negativen Effekt auf den Erwerb der SVO-Stellung des Deutschen, da die Lernenden in beiden Gruppen diese Struktur am Ende des Kurses beherrschen. Winkler geht daher davon aus, dass die Hypothese, dass der Erwerb der deutschen Wortstellungsregeln per se nicht durch Unterricht beeinflussbar ist, in dieser Eindeutigkeit nicht haltbar ist (vgl. ebd.: 319). Vielmehr sollte der Input für Lernende stärker kontrolliert werden und schon früh gezielt mehr Belege für OV-Strukturen und weniger Belege für SVO-Strukturen des Deutschen in den Unterricht eingebracht werden (vgl. ebd.: 320–322). Sie macht darauf aufmerksam, dass ihre Ergebnisse nicht den Annahmen bzw. den Voraussagen der PT widersprechen, da etwa die SVO-Stellung des Deutschen dennoch initial erworben wird (vgl. ebd.: 331).

Eine Studie, die das Desiderat aufgreift, Erwerbsstufen aus der Zweitspracherwerbsforschung mithilfe aufbereiteter und umfangreich annotierter Lernersprachkorpora mit Stufen des GER zur verknüpfen, legt in jüngster Zeit Wisniewski (2020) vor. Wisniewski untersucht mithilfe von 136 mit GER-Rating versehenen Texten auf den Sprachniveaus A2 (n=32) und B1 (n=104) aus dem MERLIN-Korpus (vgl. Abel et al. 2014) die Strukturen Inversion und Verbendstellung in Nebensätzen (vgl. Wisniewski 2020: 8–11). In ihrer Studie verwendet Wisniewski das von Pienemann definierte Emergenzkriterium von einer korrekten Realisierung in mindestens vier obligatorischen Kontexten in Kombination mit dem Emergenzkriterium von Ellis (1989), welches eine Mindestquote von 75 % korrekt realisierten Strukturen in mindestens vier obligatorischen Kontexten voraussetzt (vgl. Wisniewski 2020: 16). Zusätzlich werden Texte, in denen sich nicht genügend obligatorische Kontexte für die beiden untersuchten Strukturen finden, aus der weiteren Analyse ausgeschlossen (vgl. für eine Diskussion dieses Vorgehens Kapitel 3 und 5). Wie Jansen (2008) nimmt Wisniewski in den Blick, ob sich Belege dafür finden lassen, dass bestimmte Strukturen noch nicht emergiert sind (vgl. Wisniewski 2020: 16). Wisniewski kann, trotz der von ihr angemerkten kleinen Stichprobengröße, zeigen, dass tatsächlich statistisch signifikante Zusammenhänge zwischen den untersuchten Strukturen und den fokussierten GER-Niveaus bestehen, selbst wenn diese nicht direkt aufeinander übertragbar sind. Einerseits arbeitet die Autorin heraus, dass die Häufigkeit von realisierten Inversionsstrukturen über die beiden GER-Niveaus signifikant zunimmt und dass erst auf dem Niveau B1 eine größere Zahl von Strukturen mit Verbendstellung beobachtet werden kann (vgl. Wisniewski 2020: 29f.). Andererseits seien weder die Korrektheitsquoten der realisierten In-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

versionsstrukturen noch der Verbendstellungen ein ausschlaggebendes Kriterium für die Einordnung eines Textes auf einer bestimmten Niveaustufe durch die Bewerterinnen und Bewerter (vgl. Wisniewski 2020: 31).

2.4.3 Profilanalyse

Ein methodisches Werkzeug, das sich im Rahmen der Beschreibung und der Analyse des syntaktischen Erwerbsstandes vor allem im deutschsprachigen Raum neben den Implikationsanalysen (bzw. den Korrektheitswerten) der PT etabliert hat, ist die Profilanalyse (PA) nach Grießhaber (Grießhaber 2005; 2006; 2012; 2016).[75] Sie stellt laut Grießhaber ein Verfahren „zur sicheren Ermittlung der L2-Kenntnisse" (Grießhaber 2016: 147) von Lernenden mit Deutsch als Zweitsprache dar und wird hier vor allem im Kontext der (potenziellen) sprachlichen Förderung eingesetzt und für eine Sprachstandsbestimmung im frühen Zweitspracherwerb genutzt (vgl. Grießhaber 2012: 190; vgl. Knappik / Dirim 2018: 173). Dabei handelt es sich bei der PA aus verschiedenen Gründen, etwa der leichten Handhabbarkeit des Verfahrens für Lehrkräfte (vgl. Grießhaber 2012; Ehl et al. 2018), um ein Werkzeug, das in den letzten Jahren enorme Popularität genoss und in verschiedenen Kontexten eingesetzt wurde. Als relevant ist aus einer sprachdidaktischen Anwendungsperspektive sicherlich zu bewerten, „dass die Profilanalyse den Blick von den zahlreichen Fehlern an der sprachlichen Oberfläche auf tieferliegende Erwerbsschritte lenkt, die auch bei scheinbar wenig Hoffnung machenden Äußerungen schon erworbene Sprachmittel sichtbar machen" (Grießhaber 2012: 187). Grießhaber sieht in der PA prinzipiell ein Instrument der Zweitspracherwerbsfor-

75 Es existieren darüber hinaus zahlreiche weitere diagnostische Instrumente im Kontext von Deutsch als Zweitsprache. Im Rahmen des Projektes FörMig (vgl. Gogolin et al. 2011) sind etwa die Verfahren „Tulpenbeet" (Reich / Roth / Gantefort 2008), „Beobachtung der Schreibentwicklung" (vgl. Lengyel / Roth 2012) und „Bumerang" (Reich et al. 2009). Bei Tulpenbeet handelt es sich um eine Profilanalyse, die „für die 4.–6.Klassenstufe einsetzbar ist. Anhand einer Bildgeschichte soll ein narrativer Text geschrieben werden, der anschließend detailliert ausgewertet wird. Das verwendete Kompetenzmodell enthält sowohl Indikatoren für die Messung von Textkompetenz […] als auch für die Messung von allgemeinsprachlicher Kompetenz" (Knappik / Dirim 2018: 176). Bei dem Verfahren zur Beobachtung der Schreibentwicklung handelt es sich um „ein unterrichtsbegleitendes Beobachtungsverfahren, das die Entwicklung bildungssprachlicher Fähigkeiten im Fachunterricht fokussiert" (ebd.: 177). Bumerang ist ebenfalls eine Profilanalyse, die am Ende der Sekundarstufe 1 eingesetzt wird. Das Verfahren „besteht aus zwei Schreibaufgaben: […]. Das Profil wird auf Grundlage beider Texte erstellt, die jeweils eigens ausgewertet werden" (Knappik / Dirim: 2018: 178). Weitere Überblicksdarstellungen zu diagnostischen Instrumenten im Kontext DaZ finden sich neben Knappik / Dirim (2018) in Montanari (2018). In Bezug auf Test- und Förderdiagnostik bei erwachsenen Lernenden des Deutschen als Zweisprache gibt Feick (2018) einen umfangreichen Überblick.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

schung, mit dessen Hilfe Merkmale in der zweitsprachlichen Entwicklung herausgearbeitet werden können, die „über individuelle Verlaufsprozesse" (Grießhaber 2012: 175) hinausgehen. Während das Verfahren vor allem für Grundschulkinder in eher DaZ-orientierten Lehr-Lernzusammenhängen entwickelt und erprobt wurde (vgl. Grießhaber 2016; vgl. Goschler 2017), finden sich mittlerweile zahlreiche Operationalisierungen in weiteren Kontexten (etwa für erwachsene DaF-Lernende im chinesischen Kontext vgl. Li / Wu (2022)).

Konzeptuell handelt es sich bei Grießhabers Verfahren der Profilanalyse um eine vereinfachende Weiterentwicklung des von Clahsen (1985) und Pienemann (1998) entwickelten *rapid profiling* lernersprachlicher Äußerungen (vgl. Grießhaber 2005; 2006; 2007; u.a.), wobei Grießhaber darauf aufmerksam macht, dass die PA, die sich vor allem auf die funktional-pragmatische Kommunikationsanalyse stützt (vgl. Ehlich 2007a), „offen für unterschiedliche linguistische Konzeptionen" (Grießhaber 2012: 177) und „nicht an eine bestimmte Erwerbstheorie" (Grießhaber 2016: 147) gebunden ist. Grießhaber geht außerdem davon aus, dass die von ihm angepasste Profilanalyse sowohl auf mündliche als auch auf schriftliche Daten anwendbar ist und dass sowohl bruchstückhafte und nur schwer verständliche lernersprachliche Äußerungen als auch komplexe wissenschaftliche Textstrukturen erfasst werden können (vgl. Grießhaber 2016: 147). Tatsächlich beruht die PA im Wesentlichen auf den Entwicklungsstufen der PT (vgl. Wisniewski 2020; und dazu Tabelle 2). Unterschiede werden dagegen zu den von Diehl et al. (2000) herausgearbeiteten Erwerbsstufen deutlich:

Pienemann (1986, 1998)	**Clahsen (1985)**	**Grießhaber (Grießhaber 1999, 2012, 2016)**	**Diehl et al. (2000)**
Mündlich DaZ, DaF	Mündlich DaZ	Mündlich, schriftlich DaZ	Schriftlich DaF
		6 Integration	
		5 Insertion	
X+4 Verb Final	4 V-END	4 NS-Endstellung	V Inversion
X+3 Inversion	3 INV	3 Inversion	IV NS-Endstellung
X+2 Verb Separation	2 Sep	2 Separation	III Distanz
			II koordinierte HS
X+1 Adverb prepos.			
X Canonical order	1 Can. ord.	1 Finitum	I HS
	0	0 Bruchstücke	

Tabelle 2: Übersicht über verschiedene Modelle syntaktischer Entwicklungsstufen nach Grießhaber (2012: 176)

Wie innerhalb der PT ist eine der Grundannahmen, die der Durchführung einer PA zugrunde liegen, dass die dargestellten Entwicklungsstufen nacheinander durchlaufen werden und nicht übersprungen werden können (vgl. Grießhaber

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

2016: 153f.). Die Erwerbsstufen bilden zunehmende syntaktische Komplexität ab. In der Profilstufe werden, wie in Tabelle 2 deutlich wird, die klassischen Stufenmodelle von Clahsen und Pienemann um zwei zusätzliche Stufen erweitert: Insertion eines Nebensatzes (5) und die Integration eines erweiterten Partizipialattributes (6) (vgl. Grießhaber 2012). Diese eher stilistisch bestimmten Stufen sind vor allem in bestimmten schul- und bildungssprachlichen und in schriftlich fixierten Kontexten zu erwarten (vgl. Goschler 2017: 74). Mit der ‚Stufe 0 Bruchstücke' findet sich eine erste Stufe, die unter Umständen mit der PT-Stufe ‚Stufe 1 – Einzelwörter / Zugriff auf einzelne Lemmata' vergleichbar ist, die den anderen Stufen in der PT konzeptionell vorausgeht. Grießhaber ordnet dieser Stufe verblose oder subjektlose Äußerungen zu, die nicht analysierbar sind sowie Begrüßungs- oder Abschiedsformeln, die nie ein Verb beinhalten. Pienemann führt Adverbiale in Erstposition bei gleichzeitiger Beibehaltung der ‚kanonischen' Wortstellung als eigene Stufe, während dies bei Grießhaber nicht der Fall ist. Wird bei Grießhaber eine Adverbialphrase im Vorfeld realisiert und gleichzeitig die kanonische Wortstellung SVO beibehalten, so dass die nichtzielsprachenkonforme Wortstellung AdvSVO entsteht, ordnet Grießhaber dies je nach verwendeter Struktur der Stufe 1 ‚Finitum' oder der Stufe 2 ‚Separation' zu. Wird eine Adverbialphrase ins Vorfeld gerückt und (normgerecht) von einem Verb gefolgt, wird die minimale satzwertige Einheit der Profilstufe 3 ‚Inversion' zugeordnet.[76] Vor allem dem Erreichen der dritten Stufe schreibt Grießhaber eine besondere Bedeutung zu, „da [mit Erreichen der Stufe 3 Inversion; MS] systematisch die Satzgrenze überschritten" (Grießhaber 2016: 157) werden kann. Lernende können nun das Vorfeld „für äußerungsexterne Verweise oder für textinterne Strukturierungen" (ebd.) nutzen. Theoretisch sind Lernende jetzt in der Lage, textuelle Strukturen (d.h. zusammenhängende und aufeinander bezogene Texte) zu produzieren, die über den einzelnen Satzrahmen hinaus gehen (vgl. Grießhaber 2012: 179). Mit dem Erwerb der Inversion steht Lernenden ein Steuerungsmittel zur Verfügung, mit dessen Hilfe Hörerwissen gezielt aktiviert oder beeinflusst werden kann. Propositionale Gehalte können nun miteinander in Verbindung gebracht werden (ebd.: 180). Auf Stufe 4 sind Lernende dann in der Lage, Nebensätze mit Verbendstellung zu produzieren. Ehl et al. (2018) machen im Kontext der für die Profilanalyse vorgeschlagenen Erwerbsstufen darauf aufmerksam, dass diese „eine Erwerbsreihenfolge ohne skalierbare Abstände zwischen den Stufen darstellen" (ebd.: 1272), daher „sind die Stufen eindeutig ordinal- und nicht intervallskaliert" (ebd.). Die Autorinnen und Autoren schließen daraus, dass alle statistischen Verfahren, die im Zusammenhang mit Erwerbsstufen angewendet werden sollten, nichtparametrische Verfahren wie z.B. die Analyse von Medianen oder Rangkorrelationen sein sollten (vgl. ebd.).

76 Es existieren weitere, zum Teil leicht abweichende Erwerbsstufen im Forschungskontext DaF/DaZ, die vor allem im Bereich des kindlichen Zweitsprach-erwerbs von Tracy herausgearbeitet wurden (vgl. etwa Tracy 2007).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Bei der Durchführung einer PA werden Lerneräußerungen systematisch erfasst und komplexe, syntaktische Strukturen zunächst in minimale satzwertige Einheiten (MSE) mit einem Kern aus finitem Verb und Subjekt zerlegt (vgl. Grießhaber 2012:174; 2016: 154; zur genauen Durchführung der PA in der vorliegenden Arbeit vgl. Kapitel 3.2.4.). Jeder dieser MSE wird daraufhin die Erwerbsstufe zugeordnet, die innerhalb dieser MSE realisiert wurde. Wird eine gefundene Struktur von der Zielsprache abweichend realisiert, wird die nächsttiefere Erwerbsstufe, die korrekt realisiert wurde, zugeordnet (vgl. Beispiel a):

a) *dann wir besuchen Mdr und die Opre* (TN005, Zeitpunkt 1, Einladung nach Leipzig)

In der Äußerung von TN005 liegt durch das Adverb in initialer Position ein obligatorischer Kontext für eine Inversion vor. Diese wird nicht realisiert. Gleichzeitig liegt keine Verbseparierung vor, weshalb die Äußerung auf Erwerbsstufe 1 eingeordnet wird, da die ‚kanonische' SVO-Reihenfolge in diesem Beispiel eingehalten wird. Wenn in dieser Form alle minimalen satzwertigen Einheiten bestimmt sind, wird auf der Grundlage aller zugeordneter Strukturen das Profil für den gesamten Text bestimmt. In der von Grießhaber vorgeschlagenen PA wird eine Struktur dann als in einem Text erworben gekennzeichnet, wenn sie mindestens dreimal korrekt realisiert wurde, wobei „keine Mindestquote in einer Mindestanzahl obligatorischer Kontexte [wie in einigen PT-bezogenen Studien; MS] zugrunde gelegt" (Grießhaber 2005: 13) wird. Grießhaber macht darüber hinaus darauf aufmerksam, dass sich dieses Kriterium vor allem in schulischen Kontexten bewährt hat (vgl. Grießhaber 2016: 156). Wird in einer lernersprachlichen Äußerung eine Struktur als ‚erworben' analysiert, geht die PA davon aus, dass die Stufen mit den niedrigeren Ordnungszahlen bereits erworben wurden. Einer gesamten lernersprachlichen Äußerung wird dann die höchste im Text als erworben herausgearbeitete Struktur zugeordnet. Werden in einem Text zwei Strukturen der Erwerbsstufe 1, zwei Strukturen der Erwerbsstufe 2 und drei Strukturen der Erwerbsstufe 3 identifiziert, wird dem gesamten Text die Erwerbsstufe 3 zugeordnet.

Grießhaber geht außerdem davon aus, dass das Erreichen bestimmter (über die Wortstellung des Verbes definierter) Erwerbsstufen, zumindest indirekt, mit generellen Fertigkeiten in anderen Bereichen der L2 korreliert oder in Zusammenhang gebracht werden kann, etwa was den Wortschatz oder die Verbflexion angeht (vgl. Grießhaber 2005; 2016: 166f.).[77] Auf diesen Annahmen aufbauend könnten sich dann unter Umständen förderdiagnostische Perspektiven ergeben (vgl. Heilmann 2012). Gleichzeitig räumt Grießhaber ein, dass durch die Sprachstandsfeststellung mittels PA keine konkreten Aussagen darüber getroffen werden können, „was ein auf einer bestimmten Stufe eingeordnetes Kind mit den schon erworbenen sprachlichen Mitteln handelnd erreichen kann und was noch nicht"

77 Für einen Überblick über mögliche Zusammenhänge und eine kritische Einordnung dieser Annahmen vgl. De Carlo / Gamper (2015).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

(Grießhaber 2005: 16). Prinzipiell sieht Grießhaber die PA als interpretatives Verfahren, die einigen Spielraum zulässt, und deren Ergebnisse „nicht vorschnell in Sprachförderprogramme" (Grießhaber 2005: 16) umgesetzt werden sollten. Während in PT-bezogenen Verfahren oft Chunks vor der Analyse aus den Daten ausgeschlossen werden, werden im Rahmen der funktional-pragmatisch basierten PA „alle Lerneräußerungen erfasst und berücksichtigt" (Grießhaber 2012: 177).[78]

Aus testtheoretischer Sicht grundlegende Kritik an der PA äußern Ehl et al. (2018), die die Durchführungsobjektivität, die Auswertungsobjektivität, die Retest-Reliabilität, die Entwicklungssensitivität und die Kriteriumsvalidität der PA untersuchen (ebd.: 1263). In ihrer Evaluation der Testgüte des Verfahrens der Profilanalyse mit den Daten von insgesamt 403 ein- und mehrsprachigen Grundschulkindern (ebd.: 1272) setzen die Autorinnen und Autoren drei verschiedene Elizitationsmaterialien ein (vgl. ebd.: 1270). In ihrem Beitrag kommen die Autorinnen und Autoren zu dem Schluss, dass sowohl für unterschiedliche Elizitationsverfahren als auch für die mündliche Modalität und ebenso für die schriftliche Testung „die Durchführungsobjektivität gefährdet ist" (Ehl et al. 2018: 1277), d.h. dass jeweils nicht dieselben Erwerbsstufen ermittelt werden. Während die Re-Test-Reliabilität als ausreichend beschrieben wird, wird in der Studie von Ehl et al. nur eine geringe Interraterreliabilität (*Inter-rater reliability*), d.h., die Zuverlässigkeit mit der Bewertungen von unterschiedlichen Bewertenden übereinstimmen, erreicht, was die bisher als positiv bewertete Praktikabilität durch (linguistisch nicht ausgebildete) Lehrkräfte fraglich erscheinen lässt (vgl. Ehl et al. 2018: 1278). Zuletzt bescheinigen die Autorinnen und Autoren der PA nur eine geringe Entwicklungssensitivität in Bezug auf verschiedene Erwerbskontexte; dies hängt vor allem mit beobachteten Deckeneffekten bei den untersuchten Grundschulkindern zusammen, die dafür sorgen, dass die PA nur wenig zwischen ihnen differenzieren kann (vgl. ebd.: 1279).[79]

2.4.4 Kritik und Grenzen erwerbsstufenbezogener Ansätze

Aus den bisherigen Ausführungen zu Perspektiven auf die zweitsprachliche Entwicklung von Lernenden, die voraussagbare und linear aufeinander folgende Erwerbsstufen annehmen, lassen sich bestimmte allgemeine und spezifische

78 Für eine kritische Diskussion dieses Vorgehens innerhalb PT-basierter Studien vgl. Grießhaber (2005).

79 Als Deckeneffekte werden auf den Zweitspracherwerb bezogen Situationen bezeichnet, in denen die verwendeten Messinstrumente nicht mehr ausreichend zwischen den untersuchten Lernenden unterscheiden können, da alle oder fast alle Lernenden Werte erreichen, die gleichzeitig von den Lernenden nicht überschritten werden. Hinsichtlich der PA könnte dies dann der Fall sein, wenn alle Lernenden durchgehend auf Profilstufe 3 eingeordnet werden, aber kaum Lernende auf den Profilstufen 1 und 2 und ebenfalls keine Lernenden auf der Profilstufe 4.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Kritikpunkte herausarbeiten. Diese beziehen sich sowohl auf grundlegende Annahmen, wie Entwicklungen beschrieben werden können bzw. sollten, als auch auf konkrete methodische Verfahren, die in diesen Forschungskontexten eingesetzt werden.

Zunächst ist in der Zweitspracherwerbsforschung prinzipiell das zentrale Konstrukt ‚Erwerb' nicht immer eindeutig bestimmt bzw. bestimmbar, gleichzeitig ist nicht einwandfrei zu definieren, welches Kriterium zur Bestimmung von Erwerb zweifelsfrei herangezogen werden könnte. Selbst innerhalb eines theoretischen Frameworks wie der PT finden sich unterschiedliche Operationalisierungen zentraler Konstrukte (vgl. Nicholas et al. 2019: 391f.) und unterschiedliche Kriterien für Emergenz. Bis heute bleiben Emergenzkriterien letztlich arbiträr gesetzt und theoretisch nicht eindeutig herleitbar (vgl. für eine umfassende Zusammenfassung der Diskussion Pallotti 2007; für Annäherungen aus PT-Perspektive vgl. Clahsen 1983: 96; Lenzing et al. 2019a: 4). Weiterhin könnten unterschiedlich operationalisierte Erwerbskriterien (etwa durch unterschiedliche Mindestanwendungsquoten von Zielstrukturen in einer bestimmten Anzahl obligatorischer Kontexte) durchaus in unterschiedlichen Konzeptionen von zweitsprachlichen Erwerbsverläufen resultieren. Selbst wenn ein Emergenzkriterium eindeutig operationalisiert und für einen bestimmten Anwendungskontext umfassend begründet wird oder das Erreichen einer bestimmten Korrektheitsquote als Indikator für den Erwerb bestimmter Strukturen herangezogen wird, bleibt in PT-bezogenen Ansätzen dann unter Umständen die Frage offen, wie die weitere Entwicklung von Strukturen nach ihrer Emergenz beschrieben und erfasst werden kann (vgl. Nicholas et al. 2019: 392). Autorinnen und Autoren wie Nicholas et al. (2019) schlagen zur Operationalisierung dieser Entwicklung insgesamt den Terminus ‚die Kontrolle erhalten' (*gaining control*) (vgl. Nicholas et al. 2019: 395) vor. Dieser Prozess der Kontrollgewinnung ist ein individueller Prozess, der sich durch ein hohes Maß an intra- und interindividueller Variation auszeichnet und nicht unbedingt voraussagbar ist. Damit nähern sich die Autorinnen und Autoren einer Sicht auf zweitsprachliche Entwicklung an, die in ähnlicher Weise von Vertreterinnen und Vertretern der *Complex Dynamic Systems Theory* konzeptualisiert wird (vgl. Kapitel 2.5), selbst wenn für Nicholas et al. die zweitsprachliche Variation bei der Kontrollgewinnung weiterhin durch vorgegebene Sequenzen (innerhalb des Hypothesenraums) begrenzt ist (vgl. dazu Kapitel 2.4.2). Lenzing etwa (2015) sieht Lernersprachen zwar als dynamische Systeme, grenzt sich jedoch relativ deutlich von Ansätzen der CDST ab (für eine genauere Beschreibung dieser Ansätze vgl. das nächste Kapitel). Vielmehr geht sie davon aus, dass lernersprachliche Systeme sich sowohl durch ein hohes Maß an nicht-vorhersagbarer Variation bzw. Variabilität auszeichnen, dass diese Variation aber gleichzeitig durch die bereits erworbenen Verarbeitungsprozeduren stark limitiert wird (vgl. Lenzing et al. 2019: 396). Sie ruft das Bild eines letztlich deterministischen Chaos auf (vgl. Lenzing 2015:99). Lernersprachliche Variation ist für Lenzing zwar ein entscheidender Faktor der zweitsprachlichen Entwicklung, aber nur innerhalb der Grenzen der von der PT vorausgesagten Stufen möglich (vgl. Lenzing 2015: 117).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Grundsätzlich in Frage gestellt wird eine solche voraussagbare und letztlich lineare Perspektive auf zweitsprachliche Entwicklungen von der angesprochenen *Complex Dynamic Systems Theory*. Vertreter/-innen dieses Frameworks gehen für zweitsprachliche Entwicklungsprozesse von einer grundsätzlichen und komplexen Vernetzung aller an der zweitsprachlichen Entwicklung beteiligten (nichtlinearen) Faktoren aus, die letztlich nur eine retrospektive Analyse (und keine Vorhersagen) der dynamischen Prozesse, die in Lernenden ablaufen, ermöglicht (vgl. grundsätzlich Bot et al. 2007). Darüber hinaus wird Variabilität (bzw. Variation) als einer der zweitsprachentwicklungsspezifischen ‚Motoren' gesehen, oder anders ausgedrückt, Variabilität kann für bestimmte Entwicklungen in der Lernersprache ausschlaggebend sein (vgl. hierzu das nächste Kapitel 2.4.5). In den in diesem Kapitel geschilderten Annäherungen stellt Variabilität eher einen (sehr relevanten) Nebeneffekt der zweitsprachlichen Entwicklung dar. Die Reduzierung der zweitsprachlichen Entwicklungsprozesse auf zwar durch Variabilität geprägte, aber letztlich linear-voraussagbare Erwerbssequenzen, wird der Multikausalität und den vielen verschiedenen und komplex miteinander verbundenen Einflussfaktoren nicht gerecht (vgl. Lowie / Verspoor 2015: 74). Ebenso wie die damit einhergehende Orientierung an einer zu erreichenden L1-Norm kritisiert wird, wird die zugrundeliegende Annahme der Existenz von festen und homogenen Zielsprachen oder Erstsprachen in Frage gestellt (vgl. Larsen-Freeman 2006: 590).

Neben diesen eher grundsätzlichen Überlegungen finden sich bezüglich der verwendeten Methoden wiederholt vorgebrachte kritische Hinweise. Diese beziehen sich vor allem auf die Datensammlung und -aufbereitung sowie auf die Auswertung von Daten. Zum Teil sind diese konkret auf PT-Studien oder den Ansatz der PA bezogen, zum Teil lassen sie sich auf alle zweitspracherwerbsbezogenen Forschungskontexte übertragen. Um Entwicklungen aufzeigen zu können, sollten die erhobenen Daten entweder longitudinal sein oder aus querschnittlichen Studien mit sehr großen Datenmengen zu bestimmten (untersuchten) Phänomenen stammen, in denen die für das untersuchte Phänomen obligatorischen Kontexte (gezielt) elizitiert werden (vgl. Pienemann / Lenzing 2015: 172; für den Forschungskontext DaZ formulieren De Carlo / Gamper (2015) eine ähnliche Forderung). Hier ist zu beachten, dass Studien, die sich auf Emergenz beziehen, stets einzelne Lernende in den Blick nehmen müssen, da ein Emergenzkriterium immer nur auf Einzeltexte und nicht auf aggregierte Gruppendaten angewendet werden kann. Die Anwendung eines Emergenzkriteriums schließt folglich die Darstellung von Daten mithilfe von Gruppendurchschnitten aus, was in vielen Studienkontexten durchaus zu methodischen Problemen führen könnte (vgl. Nicholas et al. 2019: 392, aus Perspektive der CDST vgl. Lowie / Verspoor 2015). Ein solches Vorgehen findet sich jedoch in vielen der in Kapitel 2.4.2 vorgestellten Studien. Prinzipiell könnten Daten, die sich auf Gruppendurchschnitte beziehen, herangezogen werden, um die „relative Kontrolle über bestimmte Strukturen [zu betrachten; MS], nicht aber um die Emergenz dieser Strukturen zu beschreiben" (Nicholas et al. 2019: 395; Übersetzung MS). Um hier

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

für zusätzliche Transparenz zu sorgen, sollten alle Zweitspracherwerbsstudien Rohdaten und nicht nur die unter Umständen ermittelten Prozentwerte zur Verfügung stellen (vgl. Jansen 2008: 218; vgl. Kapitel 5).

In Bezug auf die bis heute herausgearbeiteten Erwerbsstufen wird von einigen Wissenschaftlerinnen und Wissenschaftlern schon relativ lange angeführt, dass diese im Grunde nicht klar genug voneinander abgrenzbar (vgl. Haberzettl 2005) oder konzeptionell zu breit angelegt sind (vgl. Meerholz-Härle / Tschirner 2001). Dies führt dazu, dass vor allem in den letzten Jahren stärker darüber nachgedacht wird, wie Variation innerhalb der Stufen stärker sichtbar gemacht werden kann (vgl. Lenzing et al. 2019) oder ob Stufen überhaupt noch aus sehr variablen Daten abgeleitet werden können (vgl. Ortega 2014: 189). Darüber hinaus können Erwerbsstufen per Definition nicht gleichzeitig erworben werden, da die niedrigere Stufe die Voraussetzung zur Verarbeitung der hierarchisch höheren Stufe darstellt (vgl. Jansen 2008); sie können durchaus in engen zeitlichen Abständen emergieren (Pienemann 1998: 248). Die Ergebnisse Pienemanns (1998) stehen damit zum Teil im Kontrast zu den Ergebnissen aus Studien wie Haberzettl (2005), Diehl et al. (2000) und Lund (2004), die ebenfalls innerhalb des PT-Frameworks durchgeführt wurden. Diese Widersprüche wurden in den betreffenden Studien nicht ausreichend oder nur am Rand thematisiert.

2.4.5 Bedeutung für DaF/DaZ

Im folgenden kurzen Abschnitt soll schlaglichtartig die Bedeutung der gerade vorgestellten Ansätze für den Fachzusammenhang DaF/DaZ vorgestellt werden. Hier liegt ein besonderer Schwerpunkt auf möglichen, aber sehr allgemeinen didaktischen Implikationen. Eine solche Darstellung greift an dieser Stelle notwendigerweise zu kurz, kann aber dennoch relevante Tendenzen aufzeigen, die dann genauer diskutiert werden könnten.

Die Bedeutung erwerbsstufenbezogener Ansätze für den sprachdidaktischen Kontext ist kaum zu überschätzen und die Vorteile solcher Perspektiven liegen relativ klar auf der Hand. Falls zweitsprachliche Erwerbs- bzw. Entwicklungsverläufe letztlich linear und weitgehend vorhersagbar verlaufen, könnten daraus klare, eindeutige und nachvollziehbare didaktische Implikationen abgeleitet werden. Gamper / Schroeder (2016) regen etwa eine didaktische Orientierung an den in diesem Kapitel vorgestellten Erwerbssequenzen an (vgl. Gamper / Schroeder 2016). Ähnlich argumentiert Thielmann (2021), der mit großem Nachdruck für die Notwendigkeit einer stärkeren Verflechtung der Zweitspracherwerbsforschung und der Sprachdidaktik vor allem im Kontext DaF/DaZ eintritt, jedoch ohne eigene empirische Daten vorzulegen und ohne auf die durchaus differenzierten Entwicklungen innerhalb der PT in den letzten Jahren weiter einzugehen. Vielversprechend sind hier korpusbasierte Ansätze, die auf annotierte und GER-Stufen zugeordnete Lernertexte zurückgreifen können. Eine solche Studie legt etwa Wisniewski (2020) vor, die die Strukturen Inversion

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

und Verbendstellung mit den Niveaustufen A2 und B1 zusammenbringt. Ebenfalls auf das Desiderat einer engeren Verknüpfung von GER-Niveaubeschreibungen und Erkenntnissen aus der Zweitspracherwerbsforschung gehen Diehl et al. (2020) ein, die sich in ihrer Argumentation auf empirische Daten aus dem DiGS-Projekt stützen (vgl. Diehl et al. 2000 und die Zusammenfassung der Studie in Kapitel 2.4.2) und anhand dieser die Möglichkeit diskutieren, dass der „grammatische Erwerbsstand eines Lernenden […] bei der Zuordnung der GER-Niveaus als Orientierungshilfe herangezogen werden" (Diehl et al. 2020: 152) könnte. Die herausgearbeiteten Ergebnisse sind aufgrund der geringen Datenmenge und dem aus der älteren DiGS-Studie übernommenen methodischen Vorgehen nur sehr bedingt aussagekräftig.

Eine Verbindung zwischen sprachlichen Entwicklungsstufen und GER-Skalen stellt demnach nach wie vor ein großes Desiderat dar, weil die GER-Niveaustufen bis heute nicht systematisch empirisch validiert sind. Dabei werden GER-Skalen oft als Abbildungen von tatsächlichen Erwerbs- bzw. Entwicklungsstufen missverstanden oder missinterpretiert (vgl. Wisniewski 2014: 403).[80] Und dies obwohl nicht unbedingt vorausgesetzt werden kann, dass „empirische Lernersprache mit den als praktikabel empfundenen Deskriptoren überhaupt beschreibbar ist" (ebd.: 403f.). Groß angelegte Projekte wie das *English Profile Project* (vgl. Hawkins / Filipović 2012) existieren für das Deutsche als L2 noch nicht.

In vielen Fällen werden die durch die PT und die PA herausgearbeiteten Erwerbsstufen direkt in didaktische Niveaubeschreibungen übernommen. So finden sich etwa in den „Niveaubeschreibungen. Deutsch als Zweitsprache für die Sekundarstufe I", des Landesamtes für Schule und Bildung des Freistaats Sachsen die von Grießhaber rekonstruierten Erwerbssequenzen im Bereich der Verbstellung (vgl. LASUB Sachsen 2019: 26).[81]

2.4.6 Zusammenfassung

In diesem Kapitel werden zunächst die Unterschiede zwischen Erst- und Zweitspracherwerb besprochen. Hier wird vor allem deutlich, dass zweitsprachliche Entwicklungsprozesse stärker von zahlreichen lernerinternen oder -externen Faktoren wie der L1, dem linguistischen Vorwissen, der Motivation, der Lernsituation etc. bestimmt zu sein scheinen. Diese Faktoren sind dabei miteinander verbunden und verändern sich im Zeitverlauf der zweitsprachlichen Entwicklung in gegenseitigen Abhängigkeiten. Um diese Entwicklungen an der sprach-

80 Thielemann (2021) unterliegt diesem Irrtum in seinem Beitrag ebenfalls oder stellt zumindest fest, dass die GER-Niveaubeschreibungen aus seiner Sicht „recht gut am tatsächlichen Spracherwerb ausgerichtet" (Thielmann 2021: 664) sind.

81 Die hier verwendeten Niveaubeschreibungen Deutsch als Zweitsprache gehen dabei vor allem auf die Arbeiten von Döll (2009; 2009a) zurück.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

lichen Oberfläche festmachen zu können, wird auf die Analyse von Lernersprache zurückgegriffen. Das Konzept von Lernersprache als eigener Sprachform wurde nicht nur, aber sehr stark, durch den Interlanguage-Ansatz von Selinker (1972) beeinflusst und in der Zweitspracherwerbsforschung als grundlegendes Konstrukt verankert. Um Lernersprachen zu untersuchen, werden oft die drei CAF-Dimensionen (*complexity*, *accuracy* und *fluency*) herangezogen. Es stehen eine Vielzahl unterschiedlicher Maße zur Verfügung, wobei in dieser Arbeit vor allem die Dimension der Komplexität im Vordergrund steht.

Gerade für schriftliche Texte sind Komplexitätsmaße relativ mühelos zu operationalisieren und anzuwenden. Deshalb finden sie ebenfalls in der deutschsprachigen Schreibforschung verstärkt Anwendung. Schreiben und Schreibentwicklung ist allerdings in besonderem Maße in institutionelle Kontexte eingebunden, was bei der Analyse von schriftlichen Daten kritisch mitbedacht werden sollte.

In Kapitel 2.4 werden zwei erwerbsstufenbasierte Ansätze zur Beschreibung der syntaktischen Entwicklung bei zweitsprachlichen Lernenden beschrieben: die *Processability Theory* und die in Ansätzen auf dieser aufbauende Profilanalyse. Die PT bietet mit der Implikationsanalyse ein präzises Beschreibungs- und Analyseinstrument zweitsprachlicher Syntaxentwicklung, während die Profilanalyse vor allem für didaktische Kontexte ein relativ niedrigschwelliges Analyseinstrument bereitstellt. Auf Grundlage zahlreicher in den letzten Jahrzehnten durchgeführten Implikationsanalysen leitet die PT sehr starke theoretische Annahmen ab, nämlich die sequenzielle Implikationalität der Erwerbsstufen und die Universalität dieser Stufen für alle Lernenden unabhängig von anderen externen Faktoren. Dabei wird innerhalb der PT mittlerweile davon ausgegangen, dass bestimmte (unter Umständen quantifizierbare) Ereignisse einem Wechsel von einer Erwerbsstufe in die nächste vorausgehen und sich die einzelnen Stufen gleichzeitig dadurch auszeichnen, dass in ihnen unterschiedliche (unter Umständen aufeinander aufbauende) Prozesse ablaufen (vgl. Nicholas et al. 2019: 396). Diese Zusammenhänge weiter und viel detaillierter zu untersuchen und herauszuarbeiten, wie sich Strukturen innerhalb des Konstrukts des Hypothesenraums entwickeln, stellt ein wichtiges Desiderat zukünftiger auf das PT-Framework bezogener Studien dar (vgl. Nicholas et al. 2019: 397).

In zahlreichen und zum Teil stark rezipierten Studien wurden die innerhalb des PT-Framework für das Deutsche herausgearbeiteten Sequenzen im Bereich der Wortstellung überprüft und im Großen und Ganzen bestätigt. In einigen Studien konnten die beschriebenen Sequenzen nicht nachgewiesen werden oder sie wurden in leicht abgewandelter Form aus den Daten rekonstruiert. Für das Deutsche wurden sowohl DaF- (Diehl et al. 2000; Bohnacker 2006; Ballestracci 2008; Jansen 2008) als auch DaZ-Lernkontexte (Czinglar 2014) untersucht. Außerdem bildeten sehr unterschiedliche Ausgangssprachen die Basis der Analysen: Diehl et al. (2000) untersuchten französischsprachige Schülerinnen und Schüler. Håkansson (2001) und Bohnacker (2006) nehmen schwedisch-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

sprachige Jugendliche und Erwachsene in den Blick. Ballestracci (2008) und Winkler (2017) gehen von italienischsprachigen Studierenden aus und Jansen (2008) untersucht englischsprachige Studierende einer australischen Universität. Obwohl die Annahmen der PT ursprünglich ausschließlich für mündliche Kontexte formuliert worden sind, werden mittlerweile sowohl mündliche als auch schriftliche Daten in longitudinalen (Pienemann 1998; Czinglar 2014) und in Querschnittsstudien (Diehl et al. 2000) sowie anhand von Implikationsanalysen und mithilfe von Korrektheitsquoten erhoben und analysiert (für einen detaillierteren Überblick vgl. Kapitel 2.4.2). In jüngerer Zeit werden Studien durchgeführt, die mithilfe aufbereiteter und offen zugänglicher lernersprachlicher Korpora Erwerbsstufen zu rekonstruieren versuchen (vgl. zum Beispiel Wisniewski (2020), die mithilfe von Daten aus dem MERLIN-Korpus erwerbsstufenspezifische Unterschiede zwischen den GER-Niveaus A2 und B1 rekonstruiert). Generell auffällig ist, dass in vielen der Studien (mit Ausnahme der großen Datensätze in Diehl et al. (2000) und Wisniewski (2020) eher wenige Lernende in den Blick genommen, aber dafür in ihrer Entwicklung sehr detailliert beschrieben werden. Während in vielen Studien die methodische Grundvoraussetzung der PT, dass von individuellen Lernendendaten ausgegangen wird, beachtet wird, gibt es vor allem im deutschsprachigen Raum zahlreiche Studien, die sich zwar auf die PT beziehen, dann aber entweder aggregierte Gruppendaten zugrunde legen oder gesamte Gruppen mit einzelnen Lernenden gleichsetzen. Dazu kommt, dass in einigen der älteren dieser Studien keinerlei deskriptive Statistiken berichtet werden und die vorgestellten Datensätze daher nur sehr begrenzt valide hinsichtlich der in den Daten vorhandenen Streuung und Variation eingeschätzt werden können. Dies ist vor allem dann problematisch, wenn etwa, wie in der DIGS-Studie, mit den Ergebnissen sehr starke (allgemeingültige) Aussagen, etwa hinsichtlich möglicher Erwerbssequenzen, verbunden werden.

Die PT ist vor allem aufgrund ihrer Fixierung auf lineare und voraussagbare Entwicklungssequenzen von Seiten der CDST (vgl. das nächste Kapitel) in die Kritik geraten. Hier sind vor allem Studien kritisch zu sehen, die sich zwar im weitesten Sinne innerhalb der *Processability Theory* verorten, in ihrer Methodik aber deutlich von Grundsätzen der PT abweichen und beispielsweise Emergenzkriterien an gemittelte Gruppendaten anlegen. Gerade in den letzten Jahren erfährt die PT eine methodische Öffnung gegenüber anderen Ansätzen der Zweitspracherwerbsforschung. Für den Forschungskontext und für die Didaktik DaF/DaZ spielt die PT nach wie vor eine bedeutende Rolle, da etwa Verfahren wie die Profilanalyse, die konzeptionell auf den Annahmen der PT beruhen, in zahlreichen Kontexten für unterschiedliche Zwecke eingesetzt werden. Mögliche Querverbindungen und Entsprechungen zwischen empirisch ermittelten Erwerbsstufen und den Niveaubeschreibungen des GER stellen ein weiterhin vielversprechendes und noch zu wenig im Fokus der Forschung stehendes Desiderat im Kontext DaF/DaZ und in der Zweitspracherwerbsforschung dar.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

2.5 *Complex Dynamic Systems Theory* im Zweitspracherwerb

In diesem Kapitel soll eine Perspektive auf zweitsprachliche Entwicklungen vorgestellt werden, die im Gegensatz zu den bisher vorgestellten Ansätzen steht – die *Complex Dynamic Systems Theory* (CDST) (vgl. Bot 2017). Fokussiert werden die Nichtlinearität und die aus der komplexen Verflechtung aller am Zweitspracherwerb beteiligten Faktoren folgende Unberechenbarkeit von zweitsprachlichen Entwicklungsprozessen. Komplexe Systeme wurden zunächst in den Naturwissenschaften und hier vor allem im Bereich der Mathematik als relevanter Untersuchungsgegenstand wahrgenommen. Mittlerweile finden sich in fast allen naturwissenschaftlichen und in immer mehr geisteswissenschaftlichen Forschungsrichtungen Ansätze, die ihre grundsätzlichen Überlegungen im Kontext der Beschreibung komplexer Systeme verorten oder zumindest Anleihen dieser Ansätze aufnehmen (vgl. Lenzing 2015: 97).[82] Die Komplexitätstheorie kann als Wissenschaft komplexer Systeme beschrieben werden (vgl. de Bot et al. 2007: 8). Dabei ist die Haupteigenschaft komplexer und dynamischer Systeme, sich im Laufe der Zeit zu verändern und von einem Zustand in andere Zustände zu wechseln, wobei sie bei ihrer Entwicklung von den ihnen zur Verfügung stehenden Ressourcen eingeschränkt werden (vgl. de Bot et al. 2007:14).

Durch ihre grundsätzliche Offenheit ist die Komplexitätstheorie auf eine Reihe von Kontexten und Untersuchungsgegenständen anwendbar und der Begriff muss eher als Überbegriff („*umbrella term*" (Hiver / Al-Hoorie 2019: 19)) verstanden werden, unter dem eine Reihe von speziellen Erkenntnisinteressen und wissenschaftlichen Foki subsumiert werden können (vgl. ebd.).[83] In diesem Zusammenhang wird von einer Metatheorie (vgl. z. B. Larsen-Freeman 2017) gesprochen, die eine spezifische Perspektive auf Veränderung (vgl. van Geert 2008: 183) einnimmt. Eine solche Metatheorie kann Ansätze in sich aufnehmen und sich die Überschneidungen in diesen Ansätzen zu Nutze machen (vgl. de Bot / Lowie / Verspoor 2007: 7). Der Begriff Komplexität lässt sich auf die lateinischen Ausdrücke *cum plectere* oder *complexus* zurückführen und meint, wörtlich übersetzt, so etwas wie ‚ineinander verwoben' oder ‚miteinander verbunden' (vgl. Hiver / Al-Hoorie 2019: 14). Komplex ist im Zusammenhang mit der wissenschaftlichen Theoriebildung nicht nur als Gegenteil von ‚einfach' zu ver-

82 Für eine ausführliche Zusammenfassung der Ursprünge und der Geschichte komplexer und dynamischer Theorien und ihrer Anwendung in verschiedenen wissenschaftlichen Disziplinen vgl. Hiver / Al-Hoorie 2019: 14–44. Die hier gegebene Zusammenfassung ist notwendigerweise sehr verkürzt und soll vor allem die Theoriebildung hinsichtlich der Anwendbarkeit im Bereich der Zweitsprachenerwerbsforschung in den Blick nehmen.

83 Hiver / Al-Hoorie (2019) nennen hier zusätzlich die Chaostheorie und nichtlineare dynamische Systeme in der Mathematik, die Synergetik in der Physik, autopoietische Systeme in der Biologie und die Philosophie des Emergentismus (vgl. ebd.: 19).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

stehen, sondern als Begriff, der die Beziehungen zwischen einzelnen Bestandteilen eines größeren Systems von Dingen beschreibt. Etwas, das als komplex beschrieben wird, ist damit „Teil eines Ganzen, das aus seinen Komponenten nicht analytisch ableitbar ist und in Kombination größere Fähigkeiten hat als die Summe seiner Teile" (Hiver / Al-Hoorie 2019: 15; [Übersetzung MS]). Der Begriff ‚dynamisch' bezieht sich darauf, dass sich diese gegenseitigen Verknüpfungen im Zeitverlauf immer wieder verändern können. Unter dem Überbegriff der Komplexitätstheorie werden folglich eine Reihe von übergreifenden, relationalen Prinzipien („relational principle"; Hiver / Al-Hoorie 2019: 21) zusammengefasst, die von komplexen Systemen grundsätzlich geteilt werden (vgl. Hiver / Al-Hoorie 2019: 20):

- **Komplexe Systeme umfassen mehrere untereinander verbundene Teile.**
  Ein komplex-dynamisches System kann als Set von Variablen bzw. von Subsystemen beschrieben werden, die miteinander interagieren und in Beziehung zueinanderstehen. Solche Systeme zeichnen sich durch eine vollständige systeminterne Vernetzung („complete interconnectedness" (de Bot et al. 2007:8)) bzw. Verschachtelung („nestedness" (Rosmawati 2020: 115)) aus.
- **Diese Teile verändern sich im Verlauf der Zeit in gemeinsamen Abhängigkeiten.**
  Dieses Netz gegenseitiger Verbindungen bringt Ergebnisse und Prozesse hervor, „die ständig wachsen, sich verändern und an neue Situationen anpassen (Hiver / Al-Hoorie 2019: 20; [Übersetzung MS]).
- **Die Prozesse dieser Veränderungen sind dynamisch und nichtlinear.**
  Mit dieser Vernetzung geht einher, dass es schwieriger (oder geradezu unmöglich) wird, genaue Vorhersagen darüber zu treffen, wie sich bestimmte Systeme im Zeitverlauf entwickeln (vgl. de Bot et al. 2007: 8). Dynamische Systeme verhalten sich unter Umständen unvorhersehbar und chaotisch, das heißt aber nicht, dass sie sich zufällig entwickeln (vgl. Lowie / Verspoor 2015: 70). Der Begriff der Nichtlinearität bezieht sich auf den Umstand, dass sich Veränderungen im Input nicht unbedingt proportional auf den beobachtbaren Output auswirken (vgl. Murakami 2014: 28).
- **Diese Prozesse führen mit der Zeit zur Emergenz von Mustern, die davor so nicht beobachtbar waren.**
  Mit Emergenz ist gemeint, dass „ein System auf einer höheren Beschreibungsebene [...] qualitativ neue Verhaltensweisen zeigt, die aus den Wechselwirkungen der Komponenten auf der niedrigeren Ebene nicht hätten vorhergesagt werden können" ((Hiver / Al-Hoorie 2019: 32; [Übersetzung MS]).

Im Folgenden werden zunächst die Grundannahmen des CDST-Ansatzes bezüglich der Zweitspracherwerbsforschung aufgezeigt. Daran schließt sich ein Überblick der Entwicklung des Ansatzes anhand relevanter empirischer Studien aus den letzten Jahren an. Ein weiterer Unterabschnitt widmet sich dann der zum Teil scharfen Kritik an diesem Ansatz. In einem weiteren Abschnitt werden

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Potenziale und Herausforderungen für den Kontext DaF/DaZ reflektiert. Wie in den vorangegangenen Kapiteln schließt dieses Kapitel mit einer kurzen Zusammenfassung.

2.5.1 Grundannahmen und theoretischer Hintergrund

Für die Anwendung bestimmter Konzepte der Komplexitätstheorie im Bereich der Zweitspracherwerbsforschung lassen sich zwei zunächst unabhängig voneinander entwickelte Ansätze ausmachen, die mittlerweile aufgrund ihrer konzeptuellen Nähe unter der Bezeichnung *Complex Dynamic Systems Theory* (CDST) zusammengefasst werden (vgl. de Bot 2017).[84] Larsen-Freeman entwickelte kurz vor der Jahrtausendwende das Label *complexity theory* (vgl. Larsen-Freeman 1998; 2002; 2007), um zweitsprachliche Entwicklungen konzeptuell in ihrer Komplexität besser fassen zu können und der großen Variabilität in lernersprachlichen Äußerungen gerecht zu werden. Für sie steht außer Frage, dass die Zweitspracherwerbsforschung sich den „messy little details" (Larsen-Freeman 2006: 614) zuwenden müsse, die für lernersprachliche Entwicklungsprozesse als charakteristisch gesehen, aber oft aus den Analysen ausgeblendet werden. Fast zeitgleich wurden ähnliche Prämissen von der niederländischen Forschungsgruppe um de Bot, Lowie und Verspoor aufgenommen, die ihre Forschungsarbeiten mit der Beschreibung *Dynamic Systems Theory* versahen. Die CDST gewann in den letzten 25 Jahren enorm an Popularität (vgl. hierzu etwa den Review-Artikel von Hiver / Al-Hoorie / Evans (2021), die fast 160 relevante Studien aus dem CDST-Kontext innerhalb der letzten 25 Jahr identifizieren (vgl. Hiver et al. 2021: 11) und den sehr kritischen Review-Artikel von Pallotti (2021); vgl. den Überblick in Bulté / Housen 2020: 208). Einen Hinweis auf die große Aufmerksamkeit, die diesem Ansatz in den letzten Jahren zukam, liefert Larsen-Freeman (2017) mit einem umfangreichen Überblick von Beiträgen aus unterschiedlichen Forschungsbereichen (vgl. ebd.: 24f.). Sie zählt, nur auf den Bereich von Forschung im Kontext von (Zweit-)Sprache bezogen, Beiträge aus den Bereichen L1-Erwerb, L2-Erwerb/Entwicklung, Neurobiologie, Kognitionswissenschaften, Sprachwandel, Sprachevolution, Sprachbewusstheit, Motivation, Diskursanalyse, Sprachpolitik, Soziolinguistik, L2-Didaktik und einige andere, nicht zuletzt den hier relevanten Bereich L2-Schreiben, auf. In diesem Zusammenhang ist der von Fogal / Verspoor herausgegebene Sammelband zu nennen (vgl. Fogal / Verspoor 2020), der sich ausschließlich der Erforschung des Schreibens in einer L2 unter Anwendung komplex-dynamischer Perspektiven widmet (vgl. dazu die ausgewählten Studien im nächsten Kapitel 2.5.2).

84 Larsen-Freeman beschreibt 2007 die grundsätzliche Komplementarität und die Notwendigkeit dialektische Beziehungen zwischen den beiden Ansätzen zu etablieren (vgl. Larsen-Freeman 2007). Heute scheinen die beiden Ansätze in den meisten Veröffentlichungen als völlig ineinander integriert verstanden zu werden. Im Fortgang dieser Arbeit wird daher die Bezeichnung CDST verwendet.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Larsen-Freeman (1997) und de Bot et al. (2007) gehen davon aus, dass die zweitsprachliche Entwicklung bei Lernenden bzw. der Zweitspracherwerb bestimmte Merkmale von komplex-dynamischen Systemen aufweist. Hier beziehen sie sich vor allem auf die Nichtlinearität und den Umstand, dass in komplex-dynamischen Systemen alle Variablen untereinander verbunden sind und miteinander in Verbindung stehen. Änderungen in einer Variablen haben notwendigerweise Auswirkungen auf das gesamte System. Sprache selbst besteht aus unterschiedlichen Systemen (etwa Morphologie, Phonologie, Syntax), die dann wieder in unterschiedliche Sprechende eingebettet sind (vgl. Lowie / Verspoor 2015: 73). Sprecherinnen und Sprecher weisen einerseits persönliche Charakteristika wie Alter, Geschlecht, kognitive Voraussetzungen etc. auf und sind andererseits in soziale Bezugssysteme eingebunden, die wiederum Teil von größeren Bildungskontexten (vgl. Byrnes 2020 zu Curricula als Subsystem von institutionellen Bildungskontexten) oder gesellschaftlichen Strukturen sind. Dabei werden Lernende sowohl durch den sie umgebenden Kontext beeinflusst als auch der Kontext selbst durch die Lernenden (vgl. Verspoor et al. 2008: 215). Die unübersichtlichen Details, die sich aus diesen vielfältigen Verbindungen ergeben, werden in der CDST zu einem relevanten Untersuchungsgegenstand (vgl. de Bot et al. 2007: 19).[85] Die CDST stellt damit einen gebrauchsbasierten (*usage-based*) Ansatz der Beschreibung zweitsprachlicher Entwicklung dar, innerhalb dessen eine Zweitsprache sich adaptiv durch Interaktion in sozialen Zusammenhängen und unter dem Einfluss vielfältiger Wechselwirkungen mit dem Kontext bzw. der Umwelt entwickelt (vgl. de Bot et al. 2007: 11; Verspoor / Behrens 2011: 27).[86] Viele der in diesem Kapitel vorgestellten Studien beschäftigen sich mit dieser Art adaptiver Anpassungen (vgl. etwa Baba / Nitta 2014; Lowie / Verspoor 2011 u.a.). Mit einer solchen Perspektive ist der Zweitspracherwerb gleichzeitig nicht mehr auf eine zu erreichende L1-Norm ausgerichtet, wie es etwa zum Teil in Interlanguage-basierten Ansätzen der Fall ist. Die Lernersprache wird als sprachliche Varietät zwischen einer Ausgangssprache und einer Zielsprache konzeptualisiert. Es gibt keinen Endzustand der zweitsprachlichen Entwicklung

85 In dieser speziellen Ausformung ihrer theoretischen Grundlagen steht die CDST teilweise im Widerspruch zum Hauptstrang „der theoretischen Linguistik[,] [die] seit etwa 1950 [...] drei Tabus eingeführt hat. Linguistische Modelle sind demnach (1) diskret (und nicht kontinuierlich) (2) linear (nichtlineare Systeme kommen nicht in Betracht) (3) deterministisch (statistische Überlegungen und Systemarchitekturen werden marginalisiert)“ (Wildgen / Plath 2005: 689).

86 In ihrem Bezug auf soziale Interaktion bzw. Kommunikation und wie sich diese im Laufe der zweitsprachlichen Entwicklung verändert, ähneln die Prämissen der CDST funktionalen Ansätzen (vgl. Bardovi-Harlig 2014: 127). Funktionelle Ansätze begreifen Zweitspracherwerbsforschung als die Erforschung der Entwicklung von alternativen Ausdrucksmöglichkeiten, die sich in der Interlanguage von Lernenden manifestieren (vgl. ebd.: 137). Für eine Einführung in funktionelle Zweitspracherwerbstheorien vgl. Bardovi-Harlig (2015).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

innerhalb einer CDST-Perspektive, vielmehr bleibt diese dauerhaften Veränderungen unterworfen (vgl. Bot et al. 2007: 19).[87]

Ein großes Potenzial der Anwendung der Komplexitätstheorie für die Zweitspracherwerbsforschung liegt darin, verschiedene bisherige Ansätze der Zweitspracherwerbsforschung in einer übergeordneten Theorie zu integrieren (vgl. de Bot et al. 2007: 19).[88] Grundsätzlich fragt die CDST folglich danach „what works, for whom, in which contexts and under what circumstances" (Hiver / Al-Hoorie 2019: 30). Die Erklärung von Kausalitäten bleibt in der CDST ein zentraler Bestandteil der wissenschaftlichen Auseinandersetzung. Zur Erforschung dieser Kausalitäten eignen sich aus Sicht der CDST vor allem Studien, die mit longitudinalen Daten bzw. Zeitreihendaten arbeiten. Solche Daten sind deshalb notwendig, weil Lernprozesse in den erwähnten Subsystemen von Sprache nicht unbedingt gleichzeitig ablaufen und Zeitreihendaten daher in besonderem Maße über die temporalen Zusammenhänge von Entwicklungen Aufschluss geben können (vgl. van Dijk / Verspoor / Lowie 2011: 59).

Eine Reihe von Vertreterinnen und Vertretern der CDST-Perspektive verbinden die Konzeptualisierung ihrer Ideen mit weitreichender Kritik an anderen Ansätzen der Zweitspracherwerbsforschung. Diese Kritik bezieht sich auf den (von CDST-Seite unterstellten) Ansatz, dass Durchschnittsdaten von Lernenden und nicht die Lernverläufe von individuellen Lernenden betrachtet und beschrieben wurden (vgl. Ortega 2014: 183). Aus Sicht des CDST-Frameworks könnte es sein, dass Studien allgemeingültige Erwerbssequenzen postuliert haben, obwohl nicht ein/-e einzige/-r Lerner/-in in ihrer bzw. seiner Entwicklung dem

87 Die Konzeption einer Interlanguage, die einen Zwischenschritt zwischen einer Ausgangssprache und einer Zielsprache darstellt, ist aus Sicht der CDST aus mehreren Gründen problematisch (vgl. Larsen-Freeman 2014: 207). Einerseits werden wieder konkrete, zielsprachliche Normen an die Analyse und die Beschreibung von Interlanguage angelegt, d. h. Interlanguage ist eben kein separates linguistisches System, sondern ein System, dass sich in einer kontinuierlichen Übergangsphase befindet, die auf eine Zielnorm hin orientiert ist. Implizit angelegt in einer solchen Perspektive ist, dass so etwas wie eine ‚perfekte Muttersprachlerin' existiert, die als Zielnorm fungieren kann. Es ist letztlich nicht genau klar, was eine solche Muttersprachlerin ausmachen soll, außer, dass sie keine Lernende ist (vgl. Larsen-Freeman 2014: 208; Han 2004: 166). Ein dritter Punkt, der sich aus der Interlanguage-Perspektive ableitet, ist letztlich, dass Erfolg im Lernen und Erwerben einer Zweitsprache sich nur innerhalb eines Vergleichs zur Zielsprache messen lässt. Die Zielsprache funktioniert als die „ultimative Messlatte des Erfolgs" (Larsen-Freemann 2014: 207 [Übersetzung MS]; vgl. Selinker 2014: 222). Das Bild einer Messlatte evoziert das Bild eines linearen L2-Erwerbs, der so nicht charakteristisch für Lernprozesse ist.

88 Kramsch (2012) geht überdies davon aus, dass diese Ansätze ein emanzipatorisches Forschungsfeld konstituieren könnten, das den Spracherwerb „aus einer kognitiven, sozialen, kulturellen, historischen und sogar ästhetischen Perspektive untersucht […]" (Kramsch 2012: 10; [Übersetzung MS]).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Durchschnittsmuster der Gruppe entsprochen hat (vgl. Lowie / Verspoor 2015: 67). Deutlicher ausgedrückt sind Verallgemeinerungen aufgrund von Durchschnittswerten nicht individualisierbar und universelle Erwerbssequenzen können nicht aus diesen abgeleitet werden. Ein weiterer methodischer Kritikpunkt, der von Seiten der CDST vorgebracht wird, bezieht sich auf den von vielen Forschenden angenommenen linearen Charakter des Zweitspracherwerbs. Aus Sicht der CDST spiegelt sich diese Perspektive dann in der Wahl bestimmter statistischer Verfahren wider. Parametrischen, statistischen Verfahren liege demnach stets die Vermutung zugrunde, dass die untersuchten Variablenbeziehungen grundsätzlich linearer Natur seien und diese Beziehungen sichtbar gemacht werden bzw. als statistisch signifikant erkannt werden müssten (vgl. Lowie / Verspoor 2015: 69). Dabei würde weiterhin davon ausgegangen, dass in Studien, die von Normalverteilungen der verwendeten Stichproben ausgehen, die Annahme zugrunde gelegt würde, dass unterschiedliche Lernende hinsichtlich bestimmter Charakteristika als homogen beschrieben werden können und daher Verallgemeinerungen zutreffend wären (vgl. Lowie / Verspoor 2015: 69). Diese Perspektive wird durch Studien aus dem CDST-Spektrum grundsätzlich in Frage gestellt (vgl. de Bot et al. 2007), da individuelle Lernverläufe, die durch unterschiedliche miteinander zusammenhängende Variablen beeinflusst werden, nicht mit interindividueller Variation zu einem Zeitpunkt gleichgesetzt werden können (vgl. Lowie / Verspoor 2015: 69). Gleichzeitig machen Vertreter/-innen der CDST darauf aufmerksam, dass die möglichen Beobachtungen direkt von den zur Verfügung stehenden Methoden abhängen (vgl. Lowie / Verspoor 2015: 84). Zur Erforschung zweitspracherwerbsbezogener Prozesse aus einer CDST-Perspektive wurde aus diesem Grund in den letzten Jahren ein spezifisches Methodenset vorgeschlagen und kontinuierlich weiterentwickelt (vgl. Verspoor / Bot / Lowie 2011; MacIntyre et al. 2017; Hiver / Al-Hoorie 2019).

Während in der Einführung in dieses Kapitel generelle Merkmale komplexer Systeme kurz angerissen wurden, geht es im weiteren Verlauf dieses Kapitels darum, die Konzepte im Rahmen der Zweitspracherwerbsforschung zu verorten und wesentliche Begrifflichkeiten zu klären. Da die Begriffsbildung der CDST-Ansätze bisher hauptsächlich in englischsprachigen Veröffentlichungen stattgefunden hat, werden an dieser Stelle jeweils die englischen Begriffe und Konzepte in Klammern mitangeführt.

– **Anfangsbedingungen eines komplex-dynamischen Systems**
  Bei der Erforschung komplex-dynamischer Systeme sind die initialen Bedingungen bzw. Anfangsbedingungen eines Systems von hoher Relevanz (*initial state*; vgl. Bot el al. 2007).[89] Kleine Unterschiede am Anfang einer Entwick-

89 Der Begriff des *initial state* spielt in der PT ebenfalls eine relevante Rolle. Lenzing (2015) geht etwa davon aus, dass dieser im Kontext zweitsprachlicher Entwicklung eine sehr frühe Erwerbsphase und nicht die Phase vor dem Erwerb einer L2 beschreibt (vgl. Lenzing 2015: 118 und dort Fußnote 3).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

lung können im weiteren Entwicklungsverlauf weitreichende Konsequenzen haben (vgl. de Bot et al. 2007: 8). Dieser Effekt wird in Anlehnung an die Begriffsbildung in der Chaostheorie vereinzelt *Butterfly Effect* genannt. In diesem Zusammenhang ist die grundsätzliche Annahme der CDST zu verstehen, dass sich komplexe Systeme durch Nichtlinearität auszeichnen (vgl. de Bot et al. 2007: 8). Es existiert demnach keine lineare Beziehung zwischen anfänglichen Störungen („perturbations"; ebd.) des Systems und der Art und Weise, wie sich die Systeme auf Dauer entwickeln (vgl. ebd.). Systeme, die anfänglich sehr ähnliche Eigenschaften aufweisen (etwa eine Gruppe von Lernenden, die sich in vielerlei Hinsicht ‚ähneln'), können sich so auf lange Sicht sehr unterschiedlich entwickeln. Wie sich ein System über einen bestimmten Zeitraum entwickelt hat und welche Rolle dabei kontextuelle Faktoren gespielt haben, kann stets nur retrospektiv beleuchtet werden. Unabhängig von ihrem Ausgangspunkt sind Systeme dauerhaft Veränderungen unterworfen, die entweder durch ihre Umwelt oder durch eine interne Selbstorganisation hervorgerufen werden. Daraus ist ableitbar, dass diese Systeme zu unterschiedlichen Zeiten auf verschiedene Weise auf den gegebenen Input reagieren werden (de Bot et al. 2007:8). Bezogen auf das Lernen einer Sprache innerhalb eines gesteuerten Kontextes bedeutet dies einerseits, dass derselbe Input bei unterschiedlichen Lernenden nicht unbedingt zu einem ähnlichen Output bei diesen führen muss (vgl. de Bot et al. 2007: 13). Andererseits leitet sich daraus ab, dass die sozialen Beziehungen innerhalb der besuchten Kursgruppe und die interagierenden Lehrpersonen, genau wie motivationale Aspekte oder psychologische Ressourcen der Lernenden, als Teil dieser initialen Bedingungen und ebenso als Teil der sich verändernden kontextuellen Bedingungen betrachtet werden müssten. Eine solch umfassende Konstruktion der interagierenden Variablen trägt dann zu einem genaueren Blick auf zweitsprachliche Entwicklungsprozesse bei (vgl. Hiver / Al-Hoorie 2019: 24).

– **Komplex-dynamische Vernetzung aller Faktoren**
In der CDST wird davon ausgegangen, dass sich komplex-dynamische Systeme durch eine grundsätzliche Offenheit gegenüber dem Kontext, in dem sie verortet sind, auszeichnen. Das bedeutet, dass alle Faktoren bzw. Variablen, die in diesen Systemen eine Rolle spielen, miteinander verbunden sind (*interconnectedness*). Komplexe Systeme zeichnen sich darüber hinaus dadurch aus, dass sie selbst wiederum mit weiteren Kontexten bzw. Systemen interagieren (vgl. Hiver / Al-Hoorie 2019: 29). In Bezug auf die zweitsprachliche Entwicklung heißt das, dass die sprachliche Kompetenz eines Lernenden zu Zeitpunkt 1 mit der Kompetenz zu den Zeitpunkten 2, 3, 4 und 5 zusammenhängt. Gleichzeitig ist diese mit der Lernmotivation und der Einstellung zum Lerngegenstand, den persönlichen Umständen, dem vorhandenen Vorwissen über weitere Sprachen oder den bekannten Lernstrategien, dem Alter und den sozial-kontextuellen Faktoren zu allen fünf Zeitpunkten verbunden (vgl. für eine grafische Darstellung dieser Zusammenhänge de Bot / Larsen-Freemann 2011: 11). Larsen-Freeman (2006) geht grundsätzlich davon aus, dass sich

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

die verschiedenen für die zweitsprachliche Entwicklung relevanten Faktoren bzw. Variablen durch unterschiedliche Beziehungen zueinander auszeichnen. Sie können kompetitiv (*competitive*), unterstützend (*supportive*) oder konditionell (*conditional*) sein (vgl. Larsen-Freeman 2006: 592–593; für weitere Erklärungen zur Operationalisierung dieser Konzepte im Rahmen der vorliegenden Studie vgl. Kapitel 3.3.3). Variablen sind dann kompetitiv, wenn sich ihre Beziehungen dadurch auszeichnen, dass ein Anstieg in den Werten der einen Variable mit einer Abnahme der Werte in einer anderen korreliert. Eine solche Beziehung ist vor allem auf die limitierten Ressourcen innerhalb eines komplex-dynamischen Systems zurückzuführen (vgl. Larsen-Freeman 2006: 593). Eine unterstützende Beziehung zwischen Variablen zeigt sich, wenn die Entwicklung sich als gemeinsamer Anstieg zweier oder mehrerer Variablen beschreiben lässt (vgl. ebd.). Larsen-Freeman bezeichnet Variablen in einer solchen Beziehung als „connected growers" (ebd.). Eine konditionale Beziehung zeichnet sich dadurch aus, dass die Entwicklungen in einer Variablen von vorangehenden Entwicklungen in einer anderen Variablen abhängen. Chan et al. (2015) bezeichnen solche Beziehungen als Vorläuferbeziehungen („precursor relationship" (Chan et al. 2015: 299)). Außerdem wird teilweise davon ausgegangen, dass Übereinstimmung (coordination) in der Entwicklung von verschiedenen Variablen einerseits als Zeichen von Stabilität (vgl. den Abschnitt „Variabilität und Stabilität") und andererseits als Hinweis auf die zunehmende Automatisierung von Prozessen gedeutet werden kann (vgl. Hou et al.: 2020: 27).

– **Entwicklung und Wachstum**
  Komplex-dynamische Systeme sind Systeme, die sich durch kontinuierlich andauernde und nichtlineare Entwicklungen auszeichnen. Entwicklung bzw. Veränderung ist in CDST-Kontexten zunächst definiert als „die Zunahme oder der Anstieg von eher fortgeschritteneren oder komplexeren Variablen und der Rückgang oder die Abnahme von weniger komplexen Variablen" (Verspoor / Schmid / Xu 2012: 242; [Übersetzung MS]). Mögliche Entwicklungen hängen zu jedem Zeitpunkt von den Ressourcen ab, die den Lernenden zur Verfügung stehen (Input, Motivation etc.) (vgl. ebd.).

– **Variabilität und Stabilität**
  Ein Schlüsselkonstrukt in Untersuchungen zu komplex-dynamischen Systemen ist Variabilität (*variability*). Generell bezeichnet Variabilität das Ergebnis der Flexibilität und der Anpassungsfähigkeit eines Systems. Im Zweitspracherwerb aus CDST-Perspektive werden mit dem Begriff in der Regel intraindividuelle lernersprachliche Unterschiede und Fluktuationen im Zeitverlauf beschrieben. Dies können zum Beispiel Unterschiede in der Verwendung von syntaktischen Strukturen wie der Wortstellung innerhalb eines Textes oder in den Daten einer Zeitreihe sein. Die CDST interessiert sich dafür, wann und wieso in verschiedenen Phasen der zweitsprachlichen Entwicklung unterschiedlich große Variabilität bzw. Stabilität beobachtet werden kann.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Gerade bei beginnenden Lernenden lassen sich große Unterschiede in der Verwendung der realisierten Strukturen beobachten, weil Lernende in dieser Phase viele verschiedene Formen und Strukturen ausprobieren, um das zu übermitteln, was sie ausdrücken wollen (vgl. van Dijk / Verspoor / Lowie 2011: 59). Dieselben Formen könnten in einem Kontext zielsprachenkonform und in einem anderen Kontext abweichend realisiert werden (vgl. Verspoor / Schmid / Xu 2012:240). Dabei könnte die Wahl für die eine oder die andere Möglichkeit relativ zufällig wirken und zusätzlich stark durch externe Faktoren wie Stress etc. beeinflusst sein (vgl. Lowie / Verspoor 2015: 76). Hohe intraindividuelle Variabilität zu Beginn eines Lernprozesses scheint oft positiv mit diesem Lernprozess zu korrelieren und könnte darauf hindeuten, dass langfristig qualitative Veränderungen stattfinden, da Variabilität hier eine wichtige Funktion erfüllt. Sie könnte so Lernenden helfen, die für ihre Kommunikation gewinnbringendsten Strategien und Strukturen auszuwählen (vgl. van Dijk / Verspoor / Lowie 2011: 58). Lowie / Verspoor (2015) sprechen von Variabilität als einem Motor für Veränderung (vgl. Lowie / Verspoor 2015: 76). Ansätze wie die PT (vgl. Kapitel 2.4) untersuchen ebenfalls Variation bzw. Variabilität, werden jedoch von der Frage geleitet, welche Informationen die Variabilität über bestimmte Entwicklungsstufen bzw. Erwerbsstufen bereithält. Es wird vielmehr systematische Variation analysiert, die nur innerhalb vorgegebener Begrenzungen bzw. innerhalb eines Hypothesenraums (vgl. Kapitel 2.4.2; Pienemann 1998; Lenzing 2015) auftritt und sich an bestimmten kontextuellen Faktoren festmachen lässt.[90] In der CDST wird, im Gegensatz dazu, davon ausgegangen, dass sich in lernersprachlichen Texten und Äußerungen freie Variabilität sichtbar machen lässt, die nicht auf Grundlage von linguistischen, kontextuellen oder psychologischen Faktoren erklärbar ist (van Dijk / Verspoor / Lowie 2011: 56–57). Prinzipiell ist es durchaus schwierig, Ursache-Wirkungs-Beziehungen zwischen Variabilität und Veränderung zu erfassen und zu interpretieren. Gleichzeitig ist es nicht unwahrscheinlich, dass diese multikausal sind (vgl. ebd.). Sprache (und hier die zu lernende Zweitsprache) befindet sich in einem konstanten Fluss (*flux*) und Unterschiede im Grad der Variabilität können Hinweise auf Entwicklungsprozesse geben (vgl. Spoelman / Verspoor 2010: 532). Dabei wird angenommen, dass Variabilität die Ursache für eine Entwicklung sein und als ein Indikator für einen bestimmten Zeitpunkt innerhalb dieser Entwicklung gesehen werden kann (vgl. Bot / Lowie / Verspoor 2007). Es wird vermutet, dass innerhalb eines Systems dann eine große Variabilität zu beobachten ist, wenn das

90 Es gab bereits in den 1970er- und 1980er-Jahren eine große Anzahl an Studien, die sich mit dem Konstrukt Variabilität beschäftigten. In der Regel stand hier die Frage danach im Vordergrund, welche Gründe sich für das Auftreten von Variabilität finden lassen und wie sich aus vorliegenden Daten Erwerbssequenzen ableiten lassen. Für einen Überblick über diese frühen Variabilitätsstudien vgl. Verspoor et al. (2008: 215–216) und Larsen-Freeman (2006: 594).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

System dabei ist, sich neu zu organisieren. Im Gegensatz dazu deutet eine geringe Variabilität auf eine relative Stabilität des beobachteten Systems hin (vgl. Verspoor et al. 2012; van Dijk / van Geert 2007; Spoelman / Verspoor 2010; van Dijk / Verspoor / Lowie 2011). Variabilität wird in der CDST demnach nicht nur als charakteristisches Hintergrundgeräusch (*noise*; vgl. Verspoor et al. 2008; Spoelman / Verspoor 2010) oder als Messfehler eines letztlich linear verlaufenden Zweitspracherwerbs, sondern als wesentliches Entwicklungsphänomen gesehen.[91] Die Untersuchung verschiedener Phasen von Variabilität innerhalb unterschiedlicher Subsysteme kann demnach Aufschluss darüber geben, wie die rekonstruierten Entwicklungsmuster der einzelnen Subsysteme zusammenwirken. Darüber hinaus könnte die Analyse von Variabilität Hinweise darauf geben, wann und warum Veränderungen in der zweitsprachlichen Entwicklung von Lernenden stattfinden. Ebenso ist Variabilität im Input für Lernende entscheidend, da unterschiedliche Strukturen helfen, verschiedene Realisierungsmöglichkeiten zu entdecken und dann aus diesen geeignete auszuwählen, um sich weiterzuentwickeln (vgl. Verspoor / Behrens 2011: 35).

Komplexe Subsysteme können in stabile Zustände gelangen. Stabilität bezeichnet den Zustand eines dynamischen Systems, in dem „kleine Störungen es qualitativ (bezogen auf Anzahl und Art der Extrema) nicht verändern" (Wildgen / Plath 2005: 691). Besonders stabile Phasen werden als Equilibrium bezeichnet (Hiver / Al-Hoorie 2019: 27–35). Bezogen auf zweitsprachliche Lernprozesse können stabile Phasen in verschiedenen Subsystemen dann erreicht bzw. beobachtet werden, wenn Lernende sich über den Gebrauch bestimmter Formen und Strukturen in unterschiedlichen Kontexten bewusst sind und sich in ihrem eigenen Gebrauch an diesen Kontexten orientieren.

Wenn dynamische Systeme im Zeitverlauf von einem speziellen Zustand, wie einem eher stabilen oder einem eher variablen Zustand, in einen anderen wechseln, ist das ebenfalls ein Wechsel von einem Attraktor (*attractor*) bzw. Attraktorzustand (*attractor state*) in einen anderen (vgl. Spoelman / Verspoor 2010: 535). Attraktoren sind Zustände, die im Laufe der Entwicklung eines dynamischen Systems wahrscheinlich erreicht und von diesen Systemen bevorzugt erreicht werden. Es existieren ebenfalls sogenannte Abstoßungszustände (*repeller states*), die Zustände markieren, die eher nicht erreicht werden (vgl. Baba / Nitta 2014: 5; Steenbeek / Jansen / van Geert 2012: 66).

91 Einer der Kritikpunkte, der von Vertreterinnen und Vertretern der CDST gegenüber anderen Ansätzen der Zweitspracherwerbsforschung immer wieder vorgebracht wird, bezieht sich darauf, dass durch die Anwendung inferenzstatistischer Verfahren die in den Daten vorhandene und relevante Variabilität als letztlich störendes ‚Hintergrundgeräusch' oder als Messfehler betrachtet wird, die den Blick auf die ‚tatsächlichen' Daten versperrt und aus diesem Grund so geglättet wird, dass sie letztlich nicht mehr sichtbar ist (vgl. van Dijk / Verspoor / Lowie 2011: 60).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Ein Erreichen dieser Zustände ist jeweils nicht voraussagbar (vgl. de Bot et al. 2007:8). De Bot et al. (2007) vergleichen diese Entwicklung entlang von und durch verschiedene Zustände mit einem über eine mit zahlreichen Schlaglöchern versehene Buckelpiste rollenden Ball (vgl. de Bot et al. 2007: 8). Dabei sind die Buckel eher Abstoßungszustände und die Schlaglöcher Attraktorenzustände. Umso tiefer ein Schlagloch ist, in das der Ball in dem Beispiel von de Bot et al. (2007) stürzt, desto mehr Energie muss aufgewendet werden, um das System wieder in einen anderen Zustand zu bringen. Bezogen auf den Erwerb einer zu erlernenden Sprache könnte beispielsweise die Zielsprache als Attraktorzustand wirken, der das lernersprachliche System anzieht (vgl. Murakami 2014: 29). In jedem Fall handelt es sich bei diesen Zuständen um temporäre Manifestationen des beobachteten Systems, die sich wieder verändern können bzw. werden. Darüber hinaus bilden verschiedene dynamische Systeme unterschiedliche dieser Attraktoren aus. Diese können sowohl zielsprachlich als auch nicht-zielsprachenkonform sein. Auf grammatischer Ebene der Lernersprachen könnte dies die Übergeneralisierung eines Pluralmorphems oder die Produktion von Inversionsstrukturen mit Verbzweitstellung sein. Im Kontext eines Sprachkurses können sich Attraktorenzustände auf die Zusammensetzung des Kurses oder bevorzugte Arbeitsgruppen beziehen. Für manche Systeme ist Chaos der wahrscheinlichste Zustand (vgl. de Bot et al. 2007: 8.). Es wird angenommen, dass ein solcher Wechsel zwischen verschiedenen Attraktoren mit einem hohen Maß an Variabilität einhergeht, die dann in dem sich verändernden System beobachtet werden kann (vgl. Spoelman / Verspoor 2010: 535). Zweitsprachliche Entwicklung zeichnet sich folglich dadurch aus, dass eine Vielzahl von komplexen Prozessen gleichzeitig und auf unterschiedlichen Zeitskalen stattfindet, die sich insgesamt zu einem kohärenten (wenn auch nicht direkt durchschaubaren) Lernprozess zusammenfügen. Dieser komplexe Lernprozess zeichnet sich durch die Emergenz neuer Muster aus (vgl. den nächsten Punkt). Wenn sich zusätzlich die neu entstehenden bzw. emergierenden Muster innerhalb des Systems durch diese Wechsel zwischen verschiedenen Attraktoren von den bisher beobachteten Mustern unterscheiden, wird von einem Phasenübergang (*phase shift*) gesprochen (Hiver / Al-Hoorie 2019: 37).

– **Emergenz und Selbstorganisation**
Wie im PT-Framework spielt der Begriff der Emergenz in der CDST eine entscheidende Rolle und wird als Schlüsselkonzept beschrieben (vgl. Baba / Nitta 2014: 4). Dabei bezeichnet Emergenz in der CDST einen (qualitativen) Phasenübergang bzw. den Übergang von einem Attraktorzustand in einen anderen. Während in der PT der Begriff der Emergenz sich auf einen Punkt in der Zeit bezieht, ab dem eine Struktur Lernenden prinzipiell zur Verfügung steht, wird der Begriff in der CDST als kontinuierliche Bewegung hin zu einer Struktur konzeptualisiert. Eine emergente Struktur ist grundsätzlich „offen und in Bewegung" (Spoelman / Verspoor 2010: 534; [Übersetzung MS]) und zeichnet sich dadurch aus, dass das gesamte System qualitative Veränderun-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

gen auf einem bisher nicht dagewesenen Komplexitätsgrad sichtbar werden lässt, wobei sich die Eigenschaft ‚qualitativ' genau auf diesen neu entstehenden Komplexitätsgrad bezieht. Das, was emergent wird, ist mehr als die Summe seiner Einzelteile (vgl. Baba / Nitta 2014: 8). Bezogen auf den Zweitspracherwerb ist daher eines der Ziele der CDST, „entscheidende Momente [die auf die Emergenz bestimmter Strukturen hindeuten; MS] in der L2-Entwicklung zu identifizieren und die Art und Weise zu erklären, wie das gesamte System in diesen Momenten funktioniert" (Baba / Nitta 2014: 5; [Übersetzung MS]). Dies kann beispielweise über die Identifikation sich verändernder Muster in den lernersprachlichen Äußerungen geschehen.

2.5.2 Empirische-methodische Entwicklung CDST-basierter Studien

In diesem Kapitel werden analog zu Kapitel 2.4.2 Studien vorgestellt, die für die Entwicklung der CDST relevant sind. Da für die Erforschung des Deutschen als L2 CDST-basierte Methoden bis heute kaum zum Einsatz kommen (vgl. aber Vyatkina et al. (2015)), behandeln die hier vorgestellten Studien in der Regel den Kontext von Englisch als L2. Außerdem wurden, mit wenigen Ausnahmen, Studien ausgewählt, die sich auf das Schreiben in einer L2 und/oder die Entwicklung syntaktischer Strukturen beziehen.

Im Anschluss an die sehr starken Aussagen, die mit dem Konzept des CDST-Ansatzes verknüpft werden, wurden in den letzten Jahren zahlreiche Studien durchgeführt, die entweder die grundsätzlich komplex-dynamische Sicht auf Sprachenlernen oder verschiedene methodische Ansätze aus dem CDST-Spektrum adaptieren. Viele dieser Studien erforschen zweitsprachliche Entwicklungsprozesse, sind aber innerhalb der Motivationsforschung angesiedelt und werden daher an dieser Stelle nicht genauer behandelt (vgl. beispielhaft den Sammelband von Dörnyei et al. 2015). Eine Ausnahme, die hier dennoch genauer aufgeführt werden soll, stellt die für die vorliegende Arbeit thematisch und methodisch relevante Studie von Pinel / Csizér (2015) dar. In dieser Studie wird mittels einer longitudinalen Clusteranalyse untersucht, wie sich im Laufe eines Semesters die Konstrukte Schreibangst (ebd.: 166), Selbstwirksamkeit (ebd.: 167) beim Schreiben und Motivation (ebd.: 168) bei 21 ungarischen Studierenden eines universitären Kurses zum wissenschaftlichen Schreiben in der L2 Englisch (ebd.: 169) verändern.[92] Pinel / Csizér sammeln Daten zu den Konstrukten anhand von Fragebögen (sechsmal im Semester) und anhand eines Essays am Ende des Semesters. Zusätzlich werden die Noten eines Schreibtests miteinbezogen, der in der Mitte des Semesters durchgeführt wird. Die Daten werden daraufhin zunächst mithilfe eines *Latent Growth Curve*-Ansatzes modelliert und mit den angesprochenen longitudinalen Clusteranalysen analysiert. Um die konstruierten Entwicklungs-

92 Vgl. zur Methode der longitudinalen Clusteranalyse und zur Anwendung in der vorliegenden Arbeit Kapitel 3.2.5.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

verläufe genauer zu untersuchen, schließen die Autorinnen Korrelations- und Variabilitätsanalysen an (vgl. Pinel / Csizér 2015: 172; Verspoor / van Dijk 2011). Die beiden Autorinnen können zeigen, dass sich die drei untersuchten Konstrukte in unterschiedlicher Weise im Laufe des Semesters verändern und sich Gruppen von Lernenden identifizieren lassen, die durch größere interindividuelle Variabilität innerhalb der untersuchten Konstrukte gekennzeichnet sind (vgl. Pinel / Csizér 2015: 185). Sie kommen zu dem Ergebnis, dass „Veränderung, oder das Fehlen derselben, mit dem Grad der Internalisierung der verschiedenen Dispositionen und Selbstwahrnehmungen zusammenhängt, d.h. dass stärker internalisierte Vorstellungen über die eigene Lernpersönlichkeit dazu neigen, Veränderungen standzuhalten, während weniger verinnerlichte Vorstellungen hinsichtlich der Selbstwahrnehmung leichter schwanken können" (ebd.: 185; [Übersetzung MS]). Damit deuten die Autorinnen an, dass bei sich neu organisierenden Systemen eine größere Variabilität zu beobachten ist. Dies ist vor allem der Fall, bevor und während die untersuchten Systeme sich verändern, während in stabilen Systemen nur wenig Variabilität zu beobachten ist.

Die Anzahl an empirischen Arbeiten, die für die vorliegende Studie explizit von Relevanz sind, weil sie sich aus CDST-Perspektive mit dem Erwerb bzw. der Entwicklung (schriftlicher) syntaktischer Strukturen bei L2-Lernenden beschäftigen, ist überschaubar. Im Folgenden werden zunächst Studien besprochen, die querschnittliche Daten analysieren. In einem zweiten Schritt werden dann longitudinale Studien präsentiert, die sich innerhalb des CDST-Kontextes verorten. Zunächst wird eine der Studien vorgestellt, die sich direkt auf die Kritik bezieht, die von Seiten der CDST an anderen Ansätzen der Zweitspracherwerbsforschung vorgebracht wird. Verspoor / Lowie / van Dijk (2008) reanalysieren Daten aus einer CDST-Perspektive, die aus einer longitudinalen Studie von Cancino / Rosansky / Schumann (1978) stammen. In der Studie von Cancino et al. (1978) werden die Daten von sechs Kindern und Jugendlichen mit Spanisch als L1 analysiert, die in einem ungesteuerten Kontext Englisch als Zweitsprache lernen. Dabei werden Negationskonstruktionen untersucht (vgl. Verspoor et al. 2008: 218), wobei ursprünglich untersucht wurde, ob sich in der zweitsprachlichen Entwicklung der untersuchten Strukturen ähnliche implikationelle und sequenziell angeordnete Erwerbsstufen wie im Erstspracherwerb rekonstruieren lassen (vgl. ebd.). In der Ausgangsstudie wurde diese Hypothese bestätigt. Verspoor et al. (2008) reanalysieren zunächst die Grafiken aus der Studie von Cancino et al. (1978) und machen darauf aufmerksam, dass sich zu Beginn der Datenerhebung nicht alle Lernenden auf derselben Stufe befinden und dass von den Lernenden zum Teil Stufen nicht durchlaufen werden (vgl. Verspoor et al. 2008: 218). Daraus ergeben sich Lernverläufe, die als nichtlinear im Vergleich zu den angenommenen Erwerbsstufen beschrieben werden können. Verspoor et al. (2008) zeigen bei ihrer Reanalyse auf, dass sich immer wieder Phasen einstellen, in denen es zu Regressphänomen in der lernersprachlichen Entwicklung kommt, die in der ursprünglichen Studie nicht im Fokus der Betrachtung standen (vgl. ebd.: 219). Anhand einer weiteren Fallstudie und mithilfe der Daten

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

eines fortgeschrittenen Englisch als Zweitsprache-Lerners mit Niederländisch als L1 diskutieren Verspoor et al. (2008) im selben Beitrag unterschiedliche Möglichkeiten, Variabilität in longitudinalen Daten sichtbar zu machen. Dabei arbeiten sie vor allem mit zuvor geglätteten Daten sowie mit min-max-Grafiken (vgl. ebd.: 220–221; vgl. für die Technik der Datenglättung mithilfe von sich bewegenden Fenstern Kapitel 3.3.3). Beide Techniken werden in der vorliegenden Studie ebenfalls angewendet (vgl. dazu Kapitel 3.3.3). Deutlich wird, dass sich selbst bei diesem fortgeschrittenen Lerner ein hohes Maß an intraindividueller Variabilität beobachten lässt und das beobachtete System nicht als stabiles System beschrieben werden kann (vgl. ebd.: 229).

Es finden sich zahlreiche Studien aus CDST-Perspektive, die Daten zu einem Punkt innerhalb der zweitsprachlichen Entwicklung, d.h. im Querschnitt, analysieren. Eine relevante Studie zum Schreiben in der L2 liefern Verspoor / Schmid / Xu (2012). In ihrer Querschnittstudie untersuchen sie, ausgehend vom *Dynamic Usage Based Approach* (ebd.: 241), insgesamt 437 schriftliche Texte von Schülerinnen und Schülern im Alter von 12 bis 14 Jahren mit der L2 Englisch und der L1 Niederländisch.[93] Zunächst ordnen acht Lehrkräfte den Texten holistisch fünf verschiedene Niveaustufen zu, die sich an den breit gefassten Kategorien Anfänger (*beginner*), Mittelstufe (*intermediate*) und Fortgeschritten (*advanced*) orientieren, aber von Verspoor et al. (2012) an dieser Stelle nicht weiter erklärt werden (vgl. ebd.: 242). Es werden insgesamt 64 Variablen anhand der Texte berechnet, die verschiedene Textmerkmale operationalisieren (etwa Satzlänge, Wortstellungsmuster, Wortschatz, Fehler, Chunks etc.). Ziel der Studie ist es, herauszuarbeiten, welche dieser Merkmale wann auftreten und am deutlichsten zwischen den angenommenen Niveaustufen diskriminieren (vgl. ebd.: 240). Die Autorinnen können zeigen, dass vor allem ‚breiter gefasste' Variablen, wie die Verwendung von Chunks oder ‚Guirauds Index' als Annährung an den von den Lernenden produzierten Wortschatz, Unterschiede zwischen den einzelnen Sprachniveaus sichtbar machen können. Textvariablen, die etwa die Länge von Sätzen oder Phrasen abbilden, entwickeln sich moderat und können nur mäßig gut zwischen den Stufen diskriminieren. Ähnlich sehen die Autorinnen die Rolle von Fehlern, da diese nicht konsistent zwischen den Niveaustufen diskriminieren können (vgl. ebd.: 255). Vielversprechend ist aus Sicht der Autorinnen die Untersuchung der Entwicklung von Chunks. Zwischen allen Niveaustufen können diese Unterschiede zeigen, wenn die Gesamtzahl aller Chunks herangezogen wird (vgl. ebd.: 251).[94]

93 Der *Dynamic Usage Based Approach* geht davon aus, dass es keine angeborene Fähigkeit für den Grammatikerwerb gibt, die die linguistische Entwicklung aller Lernender festlegt. Vielmehr wird davon ausgegangen, dass alle Lernenden aufgrund ihrer Erfahrung und aufgrund des Inputs, den sie in einer L2 erhalten, selbst Regularitäten und Muster entdecken (vgl. Verspoor / Schmid / Xu 2012: 241).

94 Dies stellt aus CDST-Perspektive ein entscheidendes Kriterium dar, da beispielsweise in vielen PT-orientierten Studien Chunks aus den Analysen ausgeschlossen werden.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Interessanterweise zeigen die Autorinnen, dass für fast alle Textvariablen die Variabilität bei beginnenden Lernenden deutlich höher als bei fortgeschrittenen ist. Die Autorinnen leiten aus ihren Ergebnissen ein erstes Entwicklungsmodell ab. Dabei gehen sie davon aus, dass Lernende zwischen Niveaustufe 1 und 2 hauptsächlich damit beschäftigt seien, Wörter zu lernen (vgl. ebd. 257), um sich dann auf der Schwelle zwischen Niveaustufe 2 und 3 mehr auf syntaktische Strukturen zu konzentrieren. Zwischen Niveaustufe 3 und 4 spielen sowohl syntaktische als auch lexikalische Strukturen eine große Rolle. Auf der Schwelle von Niveaustufe 4 zu Niveaustufe 5 scheinen die meisten syntaktischen Strukturen weit genug entwickelt und die Lernenden wenden sich wieder verstärkt lexikalischen Strukturen zu (vgl. ebd. 257). Für Verspoor et al. (2012) deuten die Ergebnisse auf eine frühe syntaktische Reorganisation und lexikalische Veränderungen zu späteren Zeitpunkten hin (vgl. ebd.: 257). Allerdings bleibt letztlich fraglich, wie die Konstruktion eines stufenbezogenen Entwicklungsmodell wie in der PT sich mit den Grundannahmen der CDST vereinbaren lässt. Dies wird von den Autor/-innen letztlich nicht ausreichend reflektiert und begründet.

Einen starken methodischen Schwerpunkt, der gleichzeitig in den größeren Kontext gebrauchsbasierter Zweitspracherwerbsforschung fällt, besitzt die Arbeit von Murakami (2014), der sich in seiner Dissertation mit interindividueller Variation bei der Entwicklung grammatischer Morpheme im Englischen als L2 (z. B. das Pluralmorphem *-s* oder das Morphem *-ing*) beschäftigt. Murakami (2014) geht aus einer komplex-dynamischen Perspektive der Frage nach, welche Rolle groß angelegte Lernerkorpora für die Zweitspracherwerbsforschung haben können. Murakami untersucht insgesamt sieben verschiedene Morpheme (vgl. Murakami 2014: 35–36) und führt insgesamt fünf Teilstudien durch, um einerseits dynamische Entwicklungsmuster herauszuarbeiten und andererseits zu klären, inwiefern ein Einfluss der verschiedenen Erstsprachen auf die zweitsprachliche Entwicklung von Morphemen besteht. Für den Kontext der hier vorliegenden Untersuchung sind nicht alle dieser kleineren Untersuchungen gleichermaßen relevant, weshalb nur die Methoden und Ergebnisse von zwei der Studien hier kurz umrissen werden. Eine der Studien implementiert ein longitudinales Forschungsdesign und wird daher etwas später in diesem Kapitel diskutiert. Das in Murakamis Untersuchungen verwendete Korpus *EF-Cambridge Open Language Database* (*EFCamDat*) umfasst im zeitlichen Kontext von Murakamis Studie die Daten von etwa 76000 Englischlernenden, die 423117 Essays verfassten (vgl. Murakami 2014: 82).[95] Murakami bildet aus diesen Daten ein Subkorpus, das mehr als 45000 Essays enthält (vgl. ebd.: 85–86), mit denen er anhand von Korrektheitsquoten die Entwicklung grammatischer Morpheme in Gruppen mit unterschiedlichen Erstsprachen rekonstruiert. Dabei wählt er

95 Das Korpus ist mittlerweile deutlich gewachsen und enthält heute (Stand Februar 2023) die Daten von mehr als 174000 Lernenden und mehr als eine Million Texte (siehe hierzu die Homepage unter: https://corpus.mml.cam.ac.uk/ [23.02.2023]; vgl. zum Korpusdesign insgesamt Huang et al. (2018); Geertzen et al. (2013).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

ein pseudolongitudinales Design mit unterschiedlichen methodischen Verfahren (Regressionsanalysen und Clusteranalysen), indem Lernende auf verschiedenen GER-Niveaustufen miteinander in Verbindung gebracht werden (vgl. ebd.: 88–89).[96] Es gelingt ihm zu zeigen, dass sich die Korrektheitsquoten in der Realisierung der untersuchten Morpheme in den verschiedenen L1-Gruppen unterscheiden (vgl. ebd.: 115).

CDST-Vertreter/-innen haben wiederholt auf die Notwendigkeit hingewiesen, Entwicklungen anhand sehr dichter Datensets und longitudinalen Zeitreihen zu untersuchen (vgl. z. B. de Bot et al. 2007; van Dijk / Verspoor / Lowie 2011; Chan / Verspoor / Vahtrick 2015). Die vielleicht erste Studie zum Schreiben in einer L2 aus CDST-Perspektive, die ebenfalls die Notwendigkeit der Analyse von longitudinalen Daten betont, ist die explorative Untersuchung von Larsen-Freeman (2006). Sie untersucht die zweitsprachliche Entwicklung von fünf Lernerinnen mit der L1 Chinesisch, die an einer amerikanischen Universität Englisch als L2 lernen (vgl. ebd.: 595). Dabei lässt Larsen-Freeman ihre Teilnehmenden insgesamt viermal dieselbe Aufgabe wiederholen. Diese besteht darin, zunächst schriftlich einen narrativen Text zu einem beliebigen Thema zu verfassen. Drei Tage nach der schriftlichen Aufgabe erzählen die Lernenden dann diesen Text erneut mündlich nach (vgl. ebd. 596). In ihrem Beitrag konzentriert sich Larsen-Freeman auf die Auswertung der schriftlichen Texte und berechnet längenbasierte Werte für unterschiedliche CAF-Dimensionen sowie Werte in Bezug auf die Komplexität des verwendeten Wortschatzes.[97] Für die Dimension der Komplexität berechnet sie die Anzahl minimaler satzwertiger Einheiten pro T-Unit. *Accuracy* wird mithilfe des Verhältnisses von fehlerfreien zu fehlerhaften T-Units und Flüssigkeit (*fluency*) mit der durchschnittlichen Anzahl von Worten pro T-Unit operationalisiert (vgl. ebd.: 597). Larsen-Freemans Ergebnisse zeigen zwei Dimensionen auf. Zunächst kann sie auf Gruppenebene zeigen, dass die Durchschnittswerte der Gruppe für alle untersuchten Variablen über die vier Zeitpunkte ansteigen (vgl. Larsen-Freeman 2006.: 599). Larsen-Freeman macht direkt darauf aufmerksam, dass diese Gruppendaten die in den einzelnen Entwicklungsverläufen vorhandene Variabilität schwerer sichtbar machen (vgl. ebd.: 598). Deshalb analysiert Larsen-Freeman in einem nächsten Schritt die individuellen Lernverläufe. Alle hier untersuchten Dimensionen von Komplexität werden mithilfe von z-Scores normalisiert und vergleichend gegenübergestellt (vgl. ebd.: 601–603). Sie kann

96 Die Essaydaten wurden über das Online-Lernprogramm *Englishtown* von *Education First* gesammelt. Die einzelnen Lektionen innerhalb dieses Programms sind an GER-Niveaustufen angelegt. Essays, die im Rahmen einer bestimmten Lektion produziert werden, ordnet Murakami den entsprechenden GER-Niveaustufen zu, d. h. es handelt sich hier nur um eine grobe Zuordnung (vgl. Murakami 2014: 81).

97 Larsen-Freeman analysiert die gesammelten Daten zusätzlich qualitativ mittels „idea units" (Larsen-Freeman 2006: 598). Auf diese Dimension der Auswertung kann in der vorliegenden Arbeit nicht weiter eingegangen werden.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

so zeigen, dass die einzelnen Entwicklungsverläufe sich alle von den vorgestellten Gruppendaten unterscheiden und sich jeweils distinkte Lernverläufe zeigen, die nicht unbedingt dem Gruppentrend folgen und zum Teil durch Regressionen und Entwicklungssprünge gekennzeichnet sind (vgl. ebd.: 604). Ihr Ansatz bleibt insgesamt rein deskriptiv und sie setzt hauptsächlich Visualisierungen ein, um ihre Ergebnisse herauszuarbeiten. Inferenzstatistische Verfahren werden von Larsen-Freeman (2006) nicht eingesetzt, was sicherlich mit dem eher explorativen Charakter der Studie insgesamt zusammenhängt.

Spoelmann / Verspoor (2010) untersuchen in einer longitudinalen Einzelfallstudie anhand einer niederländischen Lernerin individuelle Variabilität in der Entwicklung ihrer morphologischen Fertigkeiten in der L2 Finnisch. Die morphologische Entwicklung wird anhand von Korrektheitsraten und Komplexitätswerten operationalisiert, die aus einem schriftlichen Datenset extrahiert werden. In diesem Zusammenhang wird die Interaktion zwischen Korrektheit und Komplexität im zeitlichen Verlauf untersucht. Spoelman / Verspoor analysieren für ihre Studie insgesamt 54 geschriebene Texte der Lernerin, die im Laufe von drei Jahren im Rahmen eines Finnischsprachkurses an der Universität Groningen erhoben werden. Der Kurs startet für alle Lernenden ohne Vorkenntnisse im Finnischen. In Schreibaufgaben werden akademische Themen behandelt, die von der Lernerin ohne Zeitdruck als Hausaufgabe bearbeitet werden (vgl. Spoelmann / Verspoor 2010: 538). Da die Texte unterschiedlich lang sind, werden sie jeweils auf 100 Worte normalisiert bzw. es wird ein zufälliges Sample von 100 Worten aus dem Gesamttext gezogen. In diesen Samples wird die Komplexität gemessen, indem diese auf Wortebene, auf der Ebene von Nominalphrasen und auf Satzebene operationalisiert wird. Größer werdende Komplexität wird jeweils als größer werdender Anteil an den Gesamtstrukturen operationalisiert. Die Komplexität von Nominalphrasen beispielsweise berechnen Verspoor / Spoelmann (2010), indem sie die durchschnittliche Länge (in Wörtern) der Nominalphrasen in jedem Text berechnen (vgl. Verspoor / Spoelmann 2010: 539). Satzkomplexität wird durch den Durchschnitt von abhängigen satzwertigen Einheiten (*clauses*) pro Text berechnet (vgl. ebd.: 540). Die Ergebnisse werden zunächst grafisch und mit einem ähnlichen Vorgehen wie bei Verspoor et al. (2008) über (gleitende) min-max Grafiken analysiert, um die Bandbreite der beobachteten Daten sichtbar zu machen (vgl. ebd.: 538). Zusätzlich werden Progmax-Regmin Grafiken angefertigt, um mögliche Entwicklungssprünge zu identifizieren. Interaktionen zwischen Variablen werden mit verschiedenen Korrelationsanalysen berechnet. Um zu überprüfen, ob die gefundene Variabilität zufällig auftritt oder es sich tatsächlich um Entwicklungssprünge handelt, setzen die beiden Monte Carlo-Simulationen (vgl. ebd.: 539) ein.[98] Spoelmann / Verspoor (2010) gehen davon aus, dass die herausgearbeite-

98 Monte Carlo-Analysen bzw. Monte Carlo-Simulationen testen, ob gefundene Korrelationen zufällig zustande gekommen sind. Dabei wird die Variabilität mit-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

ten Entwicklungsmuster der Lernerin nichtlinear und durch einzelne Spitzen, mehrfache Rückschritte und zahlreiche Fortschritte gekennzeichnet sind. Die untersuchten Variablen stehen in komplexen Wechselwirkungen zueinander, die sich im Zeitverlauf mehrfach verändern. Die Beziehung zwischen der Komplexität von Wörtern und der Komplexität von Phrasen und Sätzen kann als unterstützend beschrieben werden. Die Werte beider Variablen steigen gemeinsam (vgl. ebd. 540). Die Maße für die Komplexität von Sätzen und die Komplexität von Nominalphrasen stehen dagegen in einem kompetitiven Verhältnis zueinander. Fallen die Werte der einen Variable, lässt sich in den Werten der anderen ein Anstieg beobachten.

Ebenfalls eine frühe Längsschnittstudie im Kontext der CDST-Theorie zum sich entwickelnden Schreiben in der L2 ist der Beitrag von Macqueen (2012). Anhand der Texte von vier L2-Englischlernenden mit unterschiedlichen Erstsprachen (Mandarin, Spanisch, Vietnamesisch), die sich auf einen Besuch einer australischen Universität vorbereiten, untersucht Macqueen die Entwicklung lexikalischer und grammatischer Muster (vgl. ebd.: 79). Dabei soll dokumentiert werden, wie sich syntagmatische Muster über die Zeit verändern (Macqueen 2012: 97). Es werden argumentative Essays und universitäre Schreibaufgaben ausgewertet und Interviews geführt.[99] Alle vier Teilnehmenden absolvieren außerdem einen individualisierten dynamischen Test, der auf bestimmte lexikalische Kombinationen aus ihrem bisherigen Schreiben abzielt. Insgesamt werden über einen Zeitraum von 9–17 Monaten 22 argumentative Essays, neun universitäre Schreibaufgaben und 25 Interviews erhoben und ausgewertet (vgl. ebd.: 79). Ein Ziel der Autorin ist es, ein Vorgehen zu entwickeln, welches qualitative Veränderungen, in der Art wie Wörter kombiniert werden, dokumentiert (vgl. Macqueen 2012: 100). Hierzu schlägt Macqueen eine Methode vor, die sie *Lexical Trail Analysis* (LTA) nennt (ebd.). Einen lexikalischen Pfad, oder *lexical trail*, definiert sie als „eine Liste aller Verwendungen eines einzelnen Lexems und seiner umgebenden Muster in chronologischer Reihenfolge" (ebd.: 100; [Übersetzung MS]).[100] Ein lexikalischer Pfad ist damit eine „detaillierte, kontextualisierte, chronologische Konkordanz" (vgl. ebd.: 102;

hilfe der Nullhypothese auf einem Konfidenzniveau von 95 % getestet. In der Studie von Spoelman / Verspoor (2010) werden die Simulationen mit insgesamt 10000 Iterationen durchgeführt (vgl. Spoelman / Verspoor 2010: 539). (vgl. z. B. Bassano / van Geert 2007 für eine genauere Beschreibung der Anwendung der Methode in einem SLA-Kontext.

99 Die Autorin weist darauf hin, dass die argumentativen Essays an der IELTS-Sprachprüfung orientiert sind; in ihnen geht es darum, eine eigene Meinung darzulegen und zu rechtfertigen (vgl. MacQueen 2012: 87). Die Schreibaufgaben werden in den universitären Kursen der Teilnehmenden unabhängig von der durchgeführten Studie gestellt.

100 Die Autorin führt als Beispiel die Entwicklung des Lexems *focus* und seinen Kollokationen an (vgl. ebd.: 101).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

[Übersetzung MS]) eines Lexems und dessen Gebrauch durch eine Lernerin oder einen Lerner. Diese bilden den Kern der Analysen und werden aus den Schreibdaten und aus den Interviewdaten abgeleitet. Aus diesen wird der dynamische Test entwickelt und die Ergebnisse dieses Tests dann wieder in die Dokumentation der lexikalischen Pfade inkorporiert, indem Strukturen als emergent oder emergierend gekennzeichnet werden (vgl. ebd. 117). Die Teilnehmenden ihrer Studie entwickeln höchst unterschiedliche lexikalische Pfade, die sich zusätzlich je nach Kontext und Funktion unterschiedlich darstellen. Macqueen kommt deshalb zu dem Schluss, dass die Entwicklung grammatisch-syntaktischer Muster einerseits und lexikalischer Muster andererseits als komplexe Interaktion analysiert werden muss (Macqueen 2012: 268). Sie schlägt ein Modell der Musterentwicklung vor, welches sich stark auf die Annahme von *zones of proximal development* (ZPD) (vgl. ebd. 270) stützt. Dieses Modell kann sich per Definition jeweils nur auf einzelne Lernende beziehen. Die Lernenden können diese ZPD durch ihren dynamisch-adaptiven Sprachgebrauch ebenfalls verändern (vgl. ebd. 278).

Eine weitere longitudinalen Studie aus der Perspektive komplex-dynamischer Systeme wird von Baba / Nitta (2014) vorgelegt, in der sie Phasenübergänge beim Schreiben untersuchen. Um einen Phasenübergang identifizieren zu können, legen Baba / Nitta (2014: 27) vier Kriterien an, die ein potenzieller Phasenübergang erfüllen muss:

- a) Es muss ein Entwicklungssprung vorliegen.
- b) Phasenübergänge müssen durch eine große Fluktuation der beobachteten Variable gekennzeichnet sein.
- c) Ein ‚kleiner Katalysator' dient letztlich als Auslöser.
- d) Es findet eine qualitative Veränderung statt.

Ziel der Studie ist, Muster in der L2-Schreibentwicklung zu identifizieren. Dabei konzentrieren sich Baba / Nitta (2014) auf eine der Dimensionen aus dem CAF-Ansatz (vgl. Kapitel 2.2), nämlich Flüssigkeit bzw. *fluency* (vgl. ebd.: 2). Flüssigkeit beim Schreiben definieren sie als geschriebene Wortzahl innerhalb einer bestimmten Zeit (vgl. ebd.: 4). Aus einer Gruppe von 22 japanischen Studierenden eines universitären Kurses zum wissenschaftlichen Schreiben in der L2 Englisch wählen sie zwei Studierende aus, bei denen die Beziehung zwischen Zeit und Veränderung in der gemessenen Schreibflüssigkeit besonders deutlich ist (vgl. ebd.: 10). Alle Lernenden produzieren innerhalb eines universitären Jahres (30 Wochen) 30 schriftliche Texte bzw. einen Text pro Woche, für den die Lernenden jeweils 10 Minuten Zeit haben. Baba / Nitta beschreiben diese Wiederholung als Kontrollparameter (vgl. ebd.)[101] Die vorgegebenen Textaufgaben

101 Für Baba / Nitta (2014) soll ein solcher Kontrollparameter vor allem dafür sorgen, dass Hintergrundvariablen so weit wie möglich kontrolliert werden. Die geschriebenen Texte bzw. die Themen/Aufgabenstellungen werden jeweils von

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

wechseln alle zwei Wochen. Die Lernenden bekommen dann jeweils drei neue Schreibanlässe vorgelegt, aus denen sie einen auswählen und dann in zwei konsekutiven Wochen einen Text zum selben Schreibanlass produzieren (vgl. ebd.: 12). Anhand der gemessenen Flüssigkeitswerte werden dann *Change Point*-Analysen durchgeführt, um Entwicklungssprünge („sudden jumps" (ebd.: 14)) zu identifizieren.[102] Im Anschluss daran untersuchen sie die identifizierten Entwicklungssprünge mithilfe von progmax-regmin Grafiken weiter (vgl. ebd.: 15). Baba / Nitta gehen in ihrer Diskussion davon aus, dass die wiederholte, zeitlich begrenzte Schreibaufgabe die Schreibflüssigkeit der beiden beobachteten Studierenden verbessert und dabei als starker Kontrollparameter funktioniert. Die herausgearbeiteten Veränderungen charakterisieren die beiden als nichtlinear (vgl. ebd.: 26). Außerdem stellen die beiden Reflexionsprozesse bei den Studierenden heraus, die sich durch das wiederholte Durchführen derselben Aufgabe beobachten lassen. Zusätzlich gehen die Autorinnen davon aus, dass sie mehrere Phasenübergänge bei den beiden Studierenden nachweisen können. Baba / Nitta reflektieren nicht, dass es in diesem Zusammenhang zu Aufgabeneffekten kommen kann. Sie führen außerdem die Verbesserung der Schreibfähigkeit auf den Einsatz der wiederholten Schreibaufgabe zurück, was vor dem Hintergrund, dass die Lernenden zusätzlich zu diesen Schreibaufgaben einen Kurs zum akademischen Schreiben besuchen, wie eine unzulässige Verkürzung wirkt. Während die Aussagekraft der Ergebnisse zusätzlich sicher durch die Konzentration auf nur eine Variable eingeschränkt ist und die Operationalisierung von Flüssigkeit unter Umständen diskussionswürdig ist, zeigen Baba / Nitta (2014) zum ersten Mal, dass es möglich ist, *Change Point*-Analysen anhand von Sprachdaten durchzuführen, wenn die vorhandene Zeitreihe dicht genug ist.

Eine sehr große Datenbasis für eine longitudinale Analyse operationalisiert Murakami (2014) in seiner bereits eingeführten Dissertation. In seiner Analyse versucht Murakami, inter- und intraindividuelle Variation bzw. Variabilität in etwa 140 000 Essays aus dem EFCAMDAT-Korpus zu entflechten (vgl. ebd.: 120). Um Muster in der zweitsprachlichen Entwicklung von grammatischen Morphemen zu identifizieren, wendet Murakami eine longitudinale Clusteranalyse an, bei der Lernende mit ähnlichen Entwicklungsverläufen zusammen gruppiert werden (vgl. ebd.). Wie bei den erwähnten querschnittlichen Ansätzen untersucht Murakami die Entwicklung anhand von Korrektheitsquoten.[103] Diese werden nun mithilfe von sich bewegenden Fenstern (*moving windows*), die sich auf 15 obligatorische Kontexte beziehen, operationalisiert (vgl. ebd.). Auf diese Weise können sehr kurze Essays ebenso analysiert werden wie längere Essays. Die

den Studierenden in ihrer Schwierigkeit bewertet, um (individuelle) Aufgabeneffekte einschätzen zu können.

102 *Change Point*-Analysen identifizieren systematische Veränderungen in einer Zeitreihe; vgl. dazu Kapitel 3.3.5.

103 An diesem Punkt gehen übrigens die klassischen Morphemstudien (z. B. Dulay / Burt 1973) ganz ähnlich vor.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

moving windows werden essay-übergreifend angelegt. Falls in einem Essay keine 15 obligatorischen Kontexte produziert wurden, erfasst das *moving window* die fehlenden Kontexte aus dem nächsten Essay der Lernenden (zum Ansatz der *moving windows* in der vorliegenden Studie vgl. Kapitel 3.3.3). Für 14144 Lernende kann Murakami longitudinale Daten rekonstruieren. Nur für etwa 1000 der Lernenden kann Murakami jedoch mehr als zehn *moving windows* anlegen (vgl. ebd.: 131). Mithilfe der Niveaustufenzuschreibung und der Erstsprachen werden diese Lernenden nun anhand ihrer Entwicklungsverläufe gruppiert. Wie in den querschnittlichen Studien operationalisiert Murakami das sprachliche Niveau anhand der Lektion innerhalb des Lernprogramms, in der die jeweiligen Essays verfasst wurden (vgl. ebd.: 133–134). Eine solche Gruppierung zeigt keine eindeutigen Ergebnisse bzw. distinkte Muster (vgl. ebd.: 135). Die Entwicklung der grammatischen Morpheme lässt sich nicht direkt auf die angesprochenen Variablen zurückführen. Um in einem nächsten Schritt Gruppen zu identifizieren, die nicht auf den genannten Metadaten beruhen, wendet Murakami zwei verschiedene Clustertechniken an: Eine regressionsbasierte Clusteranalyse (*Regression-Based Clustering*) (vgl. ebd.: 137–140), bei der bestimmte Elemente des untersuchten Entwicklungsverlaufs vorgegeben werden und eine longitudinale k-means Clusteranalyse (*k-means cluster analysis for longitudinal data*, KmL) (vgl. ebd.: 140–148). Beide Techniken zeigen bei unterschiedlichen Lernenden drei verschiedene Entwicklungsmuster in den Daten: eine insgesamt flache Entwicklung (d.h. eine gleichbleibende Korrektheitsquote), einen stetigen Anstieg der Korrektheitsquote und eine u-förmige Entwicklung dieser Quote (vgl. ebd.: 158). Darüber hinaus identifiziert Murakami in diesen drei Mustern jeweils eine sehr hohe inter- und intraindividuelle Variabilität (vgl. ebd.). Es gibt kaum Lernende, deren Entwicklung exakt den zugeordneten Mustern entspricht. Vielmehr zeichnet sich die Entwicklung der Korrektheitsquoten durch eher nichtlineare Entwicklungsverläufe aus (vgl. ebd.: 236).

Die Studie von Vyatkina / Hirschmann / Golcher (2015) stellt eine der wenigen Studien dar, die aus einer CDST-orientierten Perspektive die Entwicklung des Deutschen als L2 untersuchen. Für den Kontext der hier vorliegenden Arbeit ist die Untersuchung von Vyatkina et al. (2015) darüber hinaus relevant, da in ihr die Entwicklung syntaktischer Komplexität von Attributen (*modifier*) bei beginnenden Lernenden des Deutschen als L2 analysiert wird (vgl. ebd.: 32). Dabei nehmen die Autorinnen und Autoren eine dynamisch-gebrauchsbasierte („*dynamic usage-based*" (Vyatkina et al. 2015: 29)) Perspektive auf die zweitsprachliche Entwicklung ein. Sie analysieren sowohl Gruppendaten als auch die Daten von einzelnen Lernenden. Diese Daten stammen aus einem Subkorpus des longitudinalen KanDel-Korpus, welches zur Falko-Korpusfamilie gehört (vgl. Lüdeling et al.: 2008). Vyatkina et al. (2015) wählen für ihre Untersuchung die Daten von zwölf Studierenden mit Englisch als L1, die insgesamt 185 Texte über einen Zeitraum von vier Semestern produziert haben (vgl. Vyatkina 2012). Die Sprachdaten werden automatisiert mit den *Part-of-Speech*-Tags (POS-Tags) aus dem STTS (Schiller et al. 1999) und zusätzlich manuell mit Zielhypothesen

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

versehen (vgl. Vyatkina et al. 2015: 35–39).[104] Die Autorinnen und Autoren verwenden die POS-Tags, um Attribute in den Lernertexten zu identifizieren. Um die Veränderungen in der Menge und der Bandbreite der verwendeten Attribute aufzeigen zu können, setzen die Autorinnen und Autoren ein *generalized linear mixed model* ein.[105] Mithilfe dieser statistischen Modellierung können Vyatkina et al. (2015) zeigen, dass sich die durch verschiedene Attribute operationalisierte syntaktische Komplexität in den lernersprachlichen Äußerungen über die Studiendauer nicht wesentlich verändert. Vielmehr finden sich fast alle Arten von Attributen schon in den Texten zu Beginn der Datenerhebung (vgl. ebd.: 37). Gleichzeitig arbeiten Vyatkina et al. (2015) spezifische und sich unterscheidende Entwicklungsprofile für die einzelnen Attributstrukturen heraus. So zeigen etwa attributive Adjektive oder Relativsätze eine steigende Tendenz (vgl. 41), während Kardinalzahlen und adverbiale Adjektive im Beobachtungszeitraum in den lernerspachlichen Texten tendenziell abnehmen (vgl. ebd.). Für Präpositionalphrasen beobachten die Autorinnen und Autoren zunächst einen ansteigenden und dann einen abfallenden Trend (vgl. ebd.). Zusätzlich finden sich für alle Lernenden Zeitpunkte, an denen bestimmte Attributstrukturen nicht verwendet werden (vgl. ebd.: 43). Vyatkina et al. (2015) weisen darauf hin, dass sich trotz der zum Teil sehr klaren Entwicklungstendenzen in den Daten eine große interindividuelle Variabilität beobachten lässt, die den dynamischen Charakter der Entwicklung von syntaktischer Komplexität unterstreicht (vgl. ebd.: 43).

Eine für die vorliegende Arbeit relevante longitudinale Einzelfallstudie ist die Untersuchung von Chan / Verspoor / Vahtrick (2015), in der die Autorinnen und Autoren versuchen, Punkte in der zweitsprachlichen Entwicklung von Lernenden zu identifizieren, an denen es zu Veränderungen in den Beziehungen der Subsysteme untereinander kommt (vgl. Chan et al. 2015: 299; vgl. zu diesem Vorgehen die Operationalisierungen der Trendanalysen in der vorliegenden Arbeit in Kapitel 3.3.3). Diese Punkte sehen die Autorinnen und Autoren als Momente der Selbstorganisation des komplex-dynamischen L2-Systems (vgl. ebd.: 310). In der Studie werden mündliche und schriftliche Sprachdaten der eineiigen Zwillinge ‚Gloria' und ‚Grace' analysiert. Hier steht zusätzlich die Frage im Fokus der Untersuchung, ob sich zunehmende syntaktische Komplexität zunächst in schriftlichen oder in mündlichen Daten nachweisen lässt. Die beiden Lernerinnen sind zum Zeitpunkt der Studie 15 Jahre alt. Die Autorinnen und Autoren sammeln über einen Zeitraum von insgesamt acht Monaten etwa dreimal pro Woche Sprachdaten (vgl. ebd.: 307), wobei die Aufgaben jeweils den standardisierten TOEFL-Tests entnommen werden. Auf diese Weise erheben die

104 Part-of-speech-Tags ordnen jedes Wort eines Textes der jeweiligen Wortart zu. Dabei wird neben der Definition des Wortes der syntaktische Kontext in das Tagging einbezogen.

105 Für eine umfangreiche Einführung in diese Art von statistischer Modellierung und in *generalized linear mixed models* im Kontext der Zweitspracherwerbsforschung vgl. Gries (2021).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Autor/-innen insgesamt 100 mündliche und schriftliche Texte der beiden Lernerinnen (vgl. ebd.: 308). Aus diesen Samples werden drei syntaktische Variablen extrahiert, um syntaktische Komplexität zu messen: Die durchschnittliche Länge einer T-Unit, der Grad der Subordination und der Grad der Koordination (vgl. ebd.: 306). Diese Variablen werden automatisiert mithilfe des *automatic syntactic complexity analyzer* (Lu 2010) berechnet. Um Momente der Selbstorganisation und Veränderungen der Beziehungen zwischen den Variablen in den mündlichen und schriftlichen Daten zu identifizieren, nutzen Chan et al. (2015) eine Kombination aus t-Tests und Visualisierungen mithilfe von geglätteten Daten (vgl. ebd.: 310). Um darauf aufbauend die exakten Zeitpunkte zu bestimmen, an denen es zu Veränderungen kommt, modellieren die Autorinnen und Autoren ihre Daten mithilfe eines *Hidden Markov Model* (vgl. ebd.: 313). Für die beiden Lernerinnen werden vor allem im Bereich der schriftlichen Entwicklung unterschiedliche Muster herausgearbeitet, die auf Zeitpunkte der Selbstorganisation des Systems in Bezug auf syntaktische Komplexität hindeuten. Die Autorinnen und Autoren schließen daraus, dass sich die Länge bestimmter Lernphasen bei den beiden Lernerinnen unterscheidet und qualitativ deutlich differiert. Da sich beide Lernerinnen in Bezug auf kontextuelle Einflussfaktoren sehr ähneln, könnten diese Unterschiede letztlich mit internen Restrukturierungsprozessen zusammenhängen (vgl. ebd.: 318). Chan et al. (2015) zeigen mithilfe ihrer Analysen, dass die syntaktische Komplexität zunächst in den gesprochenen und dann in den schriftlichen Daten zunimmt (vgl. ebd.: 316).

Rosmawati (2020) untersucht in ihrer Längsschnittstudie das wissenschaftliche Schreiben einer fortgeschrittenen L2-Lernenden des Englischen. Unter Anwendung eines mehrstufigen Synchronisationsverfahrens („multilevel synchrony method" (Rosmawati 2020: 116)) versucht Rosmawati im Kontext der CDST die Verschachtelung (*nestedness*) unterschiedlicher Ebenen von syntaktischer Komplexität aufzudecken und sichtbar zu machen, wie Veränderungen auf einer Ebene andere Ebenen beeinflussen und so dynamische Wechselwirkungen zu beobachten sind (ebd.). Im Rahmen dieses Verfahrens werden „Satzbestandteile gleichzeitig zu ihrer syntaktischen Eigenschaft und zu ihrer Funktion auf der Wort-, Phrasen-, Satzteil- und Satzebene zugeordnet" (Rosmawati 2020: 116; [Übersetzung MS]). Dabei bilden Wörter die niedrigste Verschachtelungsebene. Phrasen bestehen aus Wörtern, minimale satzwertige Einheiten (*clauses*) bestehen aus Phrasen und Sätze bestehen aus minimalen satzwertigen Einheiten (vgl. ebd.). Alle Ebenen sind folglich ineinander verschachtelt. Die Ebenen der morphologischen und der lexikalischen Komplexität werden nicht miteinbezogen. Rosmawati (2020) analysiert in ihrer Untersuchung 20 schriftliche Essays einer Studentin mit Japanisch als L1, die an einer australischen Universität ein Masterprogramm absolviert. Aus jedem dieser Essays analysiert Rosmawati (2020) etwa 10 % der Textmenge, was am Ende einer Korpusgröße von etwa 4000 Wörtern entspricht (vgl. ebd.: 118). Ihre Ergebnisse legen nahe, dass vor allem die von ihr als allgemeines syntaktisches Komplexitätsmaß angenommene *Finite Verb-Token Ratio* (FVTR) über die Dauer der Datenerhebungen durch er-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

hebliche Fluktuationen bzw. Variabilität gekennzeichnet ist. Mithilfe des angewendeten Synchronisationsverfahrens kann Rosmawati zeigen, dass hohe Werte in der FVTR mit höherer Komplexität auf Phrasen- und Satzebene zusammenhängen (ebd.: 118). Die longitudinalen Daten der Lernerin zeigen, dass die Werte für Subordination auf der Ebene minimaler satzwertiger Einheiten etwa ab der Hälfte der Datenerhebung leicht abnehmen bzw. stagnieren und die Werte für phrasale Komplexität, und hier insbesondere für komplexe Nominalphrasen, stetig zunehmen (vgl. ebd.: 124–125).

2.5.3 Kritik und Grenzen von Ansätzen mit komplex-dynamischer Perspektive

Kritik erfährt der CDST-Ansatz aus verschiedenen Richtungen, vor allem von Forscherinnen und Forschern, die sich im Spektrum der PT verorten. Ein genereller Vorwurf ist, dass die Hypothesen, mit denen Vertreter/-innen der CDST arbeiten, nicht falsifizierbar seien und der gesamte Ansatz im Grunde genommen nicht den Anforderungen an eine wissenschaftliche Theorie gerecht würde (vgl. Pienemann 2015; Pallotti 2021). Vertreterinnen und Vertretern der CDST wird wiederholt vorgeworfen, letztlich bereits bekannte und relevante Konzepte der Zweitspracherwerbsforschung als ‚neue' Konzepte darzustellen, deren Erforschung in anderen Ansätzen bisher scheinbar vernachlässigt wurde. Pienemann (2007) bezieht dies konkret auf die lernersprachliche Variation bzw. Variabilität, die von der CDST als Schlüsselkonzept in den Fokus der Betrachtung gerückt wird. Er macht darauf aufmerksam, dass sich seit den 1980er-Jahren diverse Stränge der Zweitspracherwerbsforschung aus verschiedenen Perspektiven mit diesem Thema auseinandersetzen (vgl. Pienemann 2007: 43). Ganz konkret nennt er die *Processability Theory*, die lernersprachliche Variation als zentrales Konstrukt ausgemacht hat und einen Ansatz zu Erklärung dieses Phänomens bietet (vgl. Pienemann 2007: 44). Er geht im Anschluss daran davon aus, dass relevante Konzepte im Kontext der CDST noch nicht ausreichend definiert und spezifiziert sind, um deren umfassende Kritik an anderen Ansätzen zu rechtfertigen (vgl. ebd.). Ähnlich argumentieren Bulté / Housen (2020), die fünf Studien aus dem CDST-Spektrum einer kritischen Bewertung unterziehen (vgl. Bukté / Housen 2020: 212–217). Die Vorbehalte Pallottis (2021) gegenüber der CDST gehen noch weitaus tiefer. Er wirft den Vertreterinnen und Vertretern der CDST letztlich vor, Allgemeinplätze über die Komplexität von Sprachentwicklung, die als solche gar nicht kontrovers diskutiert, sondern vielmehr Gegenstand eines breiten Konsenses unter Wissenschaftlerinnen und Wissenschaftlern sind, als wissenschaftliche Fortschritte zu verkaufen. In diesem Zusammenhang kritisiert er, dass CDST-basierte Ansätze methodisch weniger leisten als andere Ansätze in der Zweitspracherwerbsforschung, da sie sich eher auf die rein deskriptive Betrachtung individueller Entwicklungsverläufe beschränken. Zusätzlich merkt Pallotti an, dass mögliche Verallgemeinerungen und überindividuelle Voraussagen in der Regel abgelehnt bzw. gar nicht erst

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

als methodische Möglichkeit oder anstrebenswertes Erkenntnisziel in Betracht gezogen werden (Pallotti 2021: 7; für eine ähnliche Argumentation hinsichtlich der angewendeten Methoden vgl. Gregg 2010). Pallottis Kritik geht so weit, dass er vorschlägt, die Benennung CDST fallen zu lassen und durch „microanalysis of variability in developmental processes" (Pallotti 2021: 11) zu ersetzen. Hier liegt gleichzeitig das Potenzial, welches Pallotti im CDST-Ansatz sieht. Die CDST lenkt den Blick auf die Untersuchung der Rolle von Variabilität in der Zweitsprachentwicklung und die Suche nach potenziellen, überindividuellen Konstanten hinsichtlich grundsätzlich variabler Entwicklungen. CDST-basierte Forschungsarbeiten müssten, ebenso wie andere theoretische Ansätze, testbare und falsifizierbare Hypothesen generieren und sich nicht auf rein deskriptive Beschreibungen individueller Entwicklungsverläufe konzentrieren.

Neben Kritik werden immer wieder Limitationen vorgebracht, die CDST-basierte Studien oft mit sich bringen. Larsen-Freeman hebt etwa an verschiedenen Stellen hervor, dass es sich bei der CDST um eine Art Metatheorie handelt, die eine Integration sehr verschiedener Ansätze erlaubt, solange bestimmte Prämissen befolgt werden (vgl. etwa Larsen-Freeman 2007; Larsen-Freeman / Cameron 2008). Zusätzlich erfordert die Konzentration auf longitudinale Daten in der Regel sehr dichte Datensets, was für viele Forschende eine enorme Herausforderung darstellt und zu einer Vernachlässigung von umfangreicheren quantitativen Studien und einer gleichzeitigen Konzentration auf Einzelfallstudien führt (vgl. Bulté / Housen 2020: 209). Gleichzeitig können CDST-Studien sich trotz der angenommenen Vernetzung von allen involvierten Faktoren in der Regel nur auf wenige, ausgewählte Variablen konzentrieren, die dann genauer analysiert werden (vgl. Lowie / Verspoor 2015: 82). Byrnes (2020) kritisiert die Auswahl der bisher untersuchten Variablen und stellt heraus, dass diese sich bisher fast immer nur auf die Satz- oder Phrasenebene bezogen und daher eher bedeutungsbezogene Text- oder Diskursperspektiven ausgeblendet wurden. Darüber hinaus wurden oft die Bildungskontexte, in denen sich Lernende befinden, trotz der grundsätzlichen Prämisse der Verschachtelung der verschiedenen Systeme und Subsysteme nicht systematisch in die Analysen miteinbezogen (vgl. Byrnes 2020: 243). Bulté / Housen (2020) machen in diesem Kontext darauf aufmerksam, dass die in einem CDST-basierten Ansatz nötige Menge an durchzuführenden Analysen und zu präsentierenden Ergebnissen schnell zu einem Überfluss an gegebenen Informationen führen könnte (vgl. Bulté / Housen 2020: 227).

2.5.4 Bedeutung für DaF/DaZ

Wie im Kapitel zur *Processability Theory* werden in diesem Kapitel kursorisch einige vor allem didaktische Implikationen angedeutet, die sich aus dem CDST-spezifischen Blick auf die zweitsprachliche Entwicklung für den Fachkontext DaF/DaZ ergeben. Dabei bleiben diese Anmerkungen, ebenso wie die Anmer-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

kungen zu PT, lediglich Annäherungen an deutlich komplexere Themen und werden daher teilweise in Kapitel 5 wiederaufgenommen und diskutiert.

Explizit CDST-basierte Ansätze haben bis jetzt kaum Eingang in den Fachzusammenhang DaF/DaZ gefunden (vgl. aber die Studie von Vyatkina et al. (2015)). Daher existieren bis jetzt keine didaktischen Annäherungen, die sich explizit an dieses Framework anlehnen.[106] Gleichzeitig verdeutlichen die Perspektiven, die in diesem Kapitel bezüglich der CDST aufgezeigt wurden, dass eine CDST-bezogene Sprachdidaktik Lösungen im Spannungsfeld zwischen einer letztlich nicht-zielorientierten Zweitsprachentwicklung und einer an einer L1-Norm ausgerichteten Sprachdidaktik anbieten müsste (vgl. Larsen-Freemann 2014: 214). Das Lernen einer Zweit- oder einer Fremdsprache findet vielmehr in der und durch die Interaktion statt und beruht auf dem Zusammenspiel einer Reihe von Faktoren. Hier passt die von der CDST vertretene Perspektive sicherlich ausgesprochen gut zu kommunikativen und handlungsorientierten Ansätzen, die das (potenzielle) Sprachhandeln der Lernenden in den Mittelpunkt stellen. Gleichzeitig müsste eine CDST-orientierte Fremd- und Zweitsprachendidaktik stark an den individuellen Lernenden und nicht an (angenommenen) Gruppen von Lernenden ausgerichtet sein. Generelle und auf die Gesamtheit aller Lernenden bezogene Aussagen, wie sie beispielsweise die stufenbezogene Perspektive der PT ermöglicht, wären kaum umsetzbar. Die CDST ist natürlich nicht die einzige Forschungsperspektive, die den Blick im Fremdsprachenlehren und -lernen auf einzelne Lernende richtet. Es existieren zahlreiche didaktische Konzepte, die mit einer CDST-orientierten Perspektive auf Zweitspracherwerb sicherlich gut zu verbinden wären, etwa Ansätze, die die Autonomie von Lernenden betonen. Insgesamt lässt sich an dieser Stelle vermuten, dass die Perspektiven, die durch die CDST auf den Zweitspracherwerb aufgemacht oder neu perspektiviert werden, sich gut mit dem verbinden könnten, was in sprachdidaktischen Lehr- und Lernkontexten im besten Fall ohnehin praktiziert wird. Einige mögliche Ansatzpunkte zukünftiger sprachdidaktischer Annäherung aus CDST-Sicht gibt Larsen-Freemann (2014). Zunächst sollte die Fremd- und Zweitsprachendidaktik darauf ausgerichtet sein, die Fähigkeiten von Lernenden in bestimmten, klar definierten Kontexten weiterzuentwickeln. Dies könnte etwa in Bezug auf bestimmte kommunikative Aufgaben oder spezielle mündliche oder schriftliche Textsorten (vgl. Fandrych / Thurmair 2011), die den Lernenden in verschiedenen Situationen begegnen könnten, geschehen (vgl. Larsen-Freeman 2014: 2014). Ein solches Vorgehen könnte Lernende dabei unterstützen, relevante sprachgebrauchsbezogene Muster in ihren tatsächlichen Gebrauchskontexten zu identifizieren und dann anzuwenden bzw. bereits bekannte sprachliche Strukturen für bestimmte Kontexte zu adaptieren. Larsen-Freeman geht außerdem davon

106 Zu einer ähnlichen Ansicht gelangt Byrnes (2020), was curriculare Vorgaben in institutionellen Bildungskontexten angeht, für den englischsprachigen Raum (vgl. Byrnes 2020: 242).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

aus, dass Sprachtests und sprachdiagnostische Instrumente nicht darauf ausgerichtet sein sollten, festzustellen, was von Lernenden bereits erworben wurde, sondern vielmehr, wie Lernfortschritte in ihrer Entwicklung sichtbar gemacht und dann bewertet werden könnten (ebd.: 216).

2.5.5 Zusammenfassung

Dieses Kapitel war der *Complex Dynamic Systems Theory* und ihrer Entwicklung in den letzten 25 Jahren gewidmet. Vertreter/-innen des CDST-Ansatzes teilen die grundsätzliche Überzeugung, dass das Lernen einer Zweitsprache als ein komplex-dynamisches System gefasst werden kann, welches sich durch nichtlineare Prozesse und eine umfassende Verschachtelung der beteiligten Systeme und Subsysteme auszeichnet. Die grundsätzliche Forschungsperspektive richtet sich daher auf zweitsprachliche Entwicklungen als Prozess und nicht auf den Erwerb zweitsprachlicher Strukturen als potenziell abgeschlossenes oder abschließbares Produkt (vgl. Lowie / Verspoor 2015: 64). (Zweit-)Sprachliche Entwicklung hat unter dieser Prämisse keinen Zielpunkt und kann daher niemals abgeschlossen sein. Zur Erforschung dieser Entwicklung werden ein eigenes Methodeninventar und eine spezifische Perspektive auf Veränderung vorgeschlagen. Aus diesem Grund eignen sich vor allem longitudinale Studien dazu, ein umfassendes Bild über die zweitsprachliche Entwicklung von Lernenden zu erhalten. Querschnittsuntersuchungen bzw. Untersuchungen, die sich auf einen bestimmten Punkt innerhalb des Lernprozesses beschränken, vermitteln im Gegensatz zu longitudinalen Studien notwendigerweise ein unvollständiges Bild von Sprachentwicklung (vgl. Lowie / Verspoor 2015: 65), können aber unter Umständen wertvolle Erkenntnisse zu sehr generellen Tendenzen der zweitsprachlichen Entwicklung liefern (z. B. ob alle Lernenden bestimmte Strukturen zu einem Zeitpunkt verwenden und wie sie das tun). Vertreter/-innen der CDST stellen sich mit Vehemenz gegen Ansätze, die Aussagen über allgemeingültige, invariante Erwerbssequenzen getroffen haben. Dabei wird vor allem die angenommene Aggregation und Analyse von Durchschnittswerten und Gruppendaten kritisiert, die aus CDST-Sicht der Analyse von individuellen Lernverläufen methodisch unterlegen ist. Problematisch werden Gruppendaten deshalb gesehen, da unter Umständen allgemeingültige Aussagen getroffen werden, die sich in den Daten von einzelnen Lernenden so nicht wiederfinden lassen (vgl. Lowie / Verspoor 2015: 67). Dabei wird die für lernersprachliche Daten charakteristische Variabilität ins Zentrum der Betrachtung gerückt. Wird die empirisch-methodische Entwicklung des CDST-Ansatzes in den Blick genommen, fallen mehrere Entwicklungslinien auf. Methodisch ist eine Tendenz zu einer sehr großen Vielfalt zu erkennen und in diesem Kontext ein (schwacher) Trend von eher deskriptiv-visuellen Verfahren zu komplexeren methodischen Verfahren. Visualisierungen spielen in CDST-Studien eine entscheidende Rolle, da zum Teil davon ausgegangen wird, dass nicht alle quantitativen Daten statistischen Testverfahren unterzogen werden sollten (vgl. Baba / Nitta 2014: 3). Gleichzeitig be-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

steht eine gewisse (durchaus so beabsichtigte) Vagheit, was den Anschluss bzw. die Anschlussfähigkeit an bisherige Theoriebildung angeht. Innerhalb der für diese Arbeit relevanten CDST-Studien lassen sich drei methodische Tendenzen erkennen:

a) Es werden Studien durchgeführt, die bereits analysierte und veröffentlichte Sprachdaten erneut analysieren (z. B. Verspoor / Lowie / van Dijk 2008).
b) Es werden Querschnittsstudien oder pseudo-longitudinale Studien mit einer großen Anzahl von Teilnehmenden durchgeführt (z. B. Verspoor / Schmid / Xu 2012; Murakami 2014).
c) Es werden longitudinale Studien durchgeführt, die in der Regel nur wenige Lernende, diese dafür zum Teil über sehr lange Zeiträume, beobachten (z. B. Larsen-Freeman (2006); Spoelman / Verspoor (2010); Macqueen (2012) Baba / Nitta (2014); Vyatkina / Hirschmann / Golcher 2015; Chan et al. (2015); Rosmawati (2020)).

Trotz oder gerade wegen ihrer in jüngster Zeit stark gestiegenen Popularität als eine Art von Metatheorie zieht die CDST starke Kritik auf sich. Diese bezieht sich vor allem darauf, dass die verwendeten Konstrukte noch nicht ausreichend operationalisiert worden seien und dass daher die durchgeführten Studien zwar eine Reihe von neuen Metaphern in den Forschungsdiskurs eingebracht hätten, in ihrer methodischen Tiefe und in den herausgearbeiteten Erkenntnissen aber hinter anderen Ansätzen der Zweitspracherwerbsforschung bis heute zurückblieben.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

3 Methodisches Vorgehen

Aufgrund der durchaus heterogenen theoretischen Ausgangslage (vgl. Kapitel 2) wurde in der vorliegenden Arbeit ein longitudinaler multimethodischer Ansatz gewählt, der sowohl theoriebasiert als auch datengeleitet verfolgt wurde. Die Analyse ist zweigeteilt. Zunächst wurden Daten auf Gruppenebene analysiert. Daran anschließend wurde mit zwei Lernenden eine Einzelfallanalyse durchgeführt, die die individuelle Entwicklung sehr ähnlicher Lernender in den Blick nimmt. Für die Analyse wurden sprachliche Strukturen ausgewählt, die zuverlässig identifiziert werden können (vgl. Petersen 2014: 80) und sich in früheren Studien als robuste Annäherungen an zweitsprachliche Entwicklungstendenzen erwiesen haben (vgl. Kapitel 2). Dabei wurden Visualisierungen verwendet, um Trends und Tendenzen, die den Daten zugrunde liegen, sichtbar und verständlich zu machen. Auf diesen grafischen Repräsentationen der Daten aufbauend wurden dann ergänzende statistische Verfahren angewendet. Statistische Analysen wurden jeweils mit deskriptiven Daten vorbereitet und dann mithilfe (nicht-)parametrischer Verfahren durchgeführt.[107] Diese Verfahren wurden durch robuste Bootstrapping-Analysen ergänzt.[108] Bei Verfahren, die

107 Nicht-parametrische Tests sind wie ihre parametrischen Gegenstücke statistische Verfahren, mit denen Hypothesen überprüft werden können. Im Gegensatz zu parametrischen zeichnen sich nicht-parametrische Verfahren dadurch aus, dass sie „keine spezielle Verteilungsannahme [etwa eine Normalverteilung; MS] für die Grundgesamtheit und kein kardinales Meßniveau [sic] der Daten" (Büning / Trenkler 1994: 2) erfordern. Liegt keine Normalverteilung vor, messen nicht-parametrische Verfahren effizienter. Gleichzeitig „stellt sich nicht in dem Maße wie bei parametrischen Verfahren [das Robustheitsproblem], weil die zur Anwendung von nichtparametrischen Verfahren erforderlichen (schwachen) Annahmen in der Regel erfüllt sind" (ebd.). Die den nicht-parametrischen Tests jeweils zugrundeliegenden Annahmen werden in den jeweiligen Kapiteln gesondert vorgestellt. Für einen Vergleich einiger ausgewählter parametrischer und nicht-parametrischer Tests vgl. Reuschenbach (2009). Als Alternative zu nicht-parametrischen Verfahren werden immer häufiger sogenannte ‚robuste Methoden' eingesetzt (vgl. für eine Einführung und eine Diskussion von Vor- und Nachteilen von nicht-parametrischen Verfahren im Vergleich zu robusten Methoden Field (2018: 265–268)). Zu den robusten Methoden zählt unter anderem das sogenannte Bootstrapping, das ebenfalls in der vorliegenden Arbeit eingesetzt wird.

108 Bootstrapping ist eine Resampling-Technik, die es über die Ziehung von vielen Stichproben aus einer bekannten Stichprobe ermöglicht, Informationen „über eine ansonsten „unbekannte" Grundgesamtheit" (Backhaus et al. 2018: 357) zu erhalten. Oft werden aus einer Stichprobe 1000 oder 10000 Bootstraps gezogen. Die bekannte Stichprobe wird damit im Grunde wie die Grundgesamtheit behandelt, aus der neue Stichproben gezogen werden (vgl. Field 2018: 1009). Für jede Stichprobe werden Lageparameter (wie das arithmetische Mittel oder die Standardabweichung) berechnet, mit deren Hilfe dann Konfidenzintervalle be-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Unterschiede testen und bei Korrelationsanalysen wurden jeweils Effektstärken angegeben (vgl. Plonsky / Oswald 2014; vgl. Larson-Hall / Plonsky 2015).[109] Bei allen im Rahmen dieser Studie durchgeführten Korrelationsanalysen wurde ein α-Level von 0,05 angesetzt. Zusätzlich werden in der Regel Konfidenzintervalle angegeben (vgl. Larson-Hall / Plonsky 2015). Die folgenden drei Kapitel stellen die Grundlagen der Studie und den Studienaufbau (vgl. Kapitel 3.1), das methodische Vorgehen hinsichtlich der Analysen auf Gruppenebene (vgl. Kapitel 3.2) und der Einzelfallanalysen (vgl. Kapitel 3.3) vor.

3.1 Methodische Grundlagen und Datenbasis

In den nun folgenden Ausführungen zu den methodischen Grundlagen der Studie werden zunächst die Ausgangspunkte der Studie geschildert. Dabei werden einerseits Zugangsvoraussetzungen vorgestellt, die für eine Teilnahme eine Rolle spielten (vgl. Kapitel 3.1.1). Und andererseits wird gezeigt, wie sich aufgrund dieser Zugangsvoraussetzungen die Gruppe der Teilnehmenden der Gesamtstudie zusammensetzte (vgl. Kapitel 3.1.2). Danach folgt eine Systematisierung des zeitlichen Ablaufs (vgl. Kapitel 3.1.3 und 3.1.4) und Erklärungen zur experimentellen Manipulation (vgl. Kapitel 3.1.5), die dem Forschungsprojekt zugrunde lag. In einem weiteren Abschnitt werden Limitationen der Studie reflektiert (vgl. Kapitel 3.1.6). Hier wird zunächst die Position des Forschenden innerhalb des Studienzusammenhangs geklärt (vgl. Kapitel 3.1.6.1) und die Zugänglich-

rechnet oder Signifikanztests durchgeführt werden können (vgl. ebd.). Efron / Tibshirani (1994: 5) führen den Begriff des Bootstrapping in der Statistik auf den englischen Ausdruck *to pull oneself up by one's bootstraps* zurück, den Backhaus et al. (2018: 357) mit „an den eigenen Haaren aus dem Sumpf zieh[en]" wiedergeben. Zur Technik des Bootstrapping vgl. grundlegend Efron / Tibshirani (1994) und Field (2018: 264–268).

109 Effektstärken bzw. Effektgrößen sind standardisierte, ergänzende Angaben zu statistischen Signifikanztests, die es erlauben, Ergebnisse unterschiedlicher Studien miteinander zu vergleichen (vgl. Plonsky / Oswald 2014: 880). Sie sind vor allem dann nützlich, wenn Daten aus verschiedenen Studien zu Metaanalysen kombiniert werden sollen, um ein tieferes Verständnis bestimmter Phänomene zu erhalten. Metaanalysen beruhen weitgehend auf dem korrekten und für alle Tests durchgehenden Berichten von Effektstärken (vgl. Larson-Hall / Plonsky 2015: 132f.). Für eine Einführung und eine zusammenfassende Diskussion, warum *p*-Werte durch Effektstärken ergänzt werden sollten vgl. Plonsky / Oswald (2014: 879–881). In der vorliegenden Studie werden Effektstärken in ihrer Größe nach der Klassifizierung nach Cohen (1998) eingeschätzt. Dies bedeutet, dass für Cohens $d = 0.2$ eine kleine, 0,5 eine mittlere und 0,8 eine große Effektstärke darstellt. Für r wird 0,1 als kleine, 0,3 als mittlere und 0,5 als große Effektstärke gesehen. Plonsky / Oswald (2014) evaluieren diese Effektstärken im Kontext von Zweitspracherwerbsstudien und schlagen eine leicht abweichende Klassifizierung vor.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

keit der Daten besprochen (vgl. Kapitel 3.1.6.2). In einem weiteren Schritt wird die Gruppe der Abbrecherinnen und der Abbrecher mit den Teilnehmenden verglichen (vgl. Kapitel 3.1.6.3), die die Studie weiter durchgeführt haben.[110] Anschließend an dieses einführende Kapitel wird das methodische Vorgehen hinsichtlich der Forschungsfragen und der zur Beantwortung dieser Fragen gebildeten Teilstichproben spezifiziert (vgl. Kapitel 3.2 und Kapitel 3.2).

3.1.1 Rekrutierung der Teilnehmenden und Zugangsvoraussetzungen

Insgesamt wurden zu fünf Terminen etwa 250 potenzielle Teilnehmende, die fast ausschließlich als Geflüchtete nach Leipzig gekommen waren, an das Max-Planck-Institut für Kognitions- und Neurowissenschaften in Leipzig eingeladen und dort über die geplante Studie informiert.[111] Alle Informationstreffen wurden auf Arabisch durchgeführt und die relevanten Dokumente und Formulare davor ins Arabische übersetzt. Zusätzlich zum Studienarzt des Max-Planck-Instituts nahm ein arabischsprachiger Arzt an den Informationsveranstaltungen teil, der den Auswahlprozess begleitete und den Interessierten bei Fragen zur Verfügung stand. Die begleitende ärztliche Aufklärung ist für alle Studien verpflichtend, die am Max-Planck-Institut für Kognitions- und Neurowissenschaften in Leipzig durchgeführt werden. Einer der wesentlichen Punkte der ärztlichen Aufklärung ist die Erklärung der Funktionsweise der nicht-invasiven MRT-Scans des Gehirns, die im Rahmen der Studie durchgeführt wurden.

Um an der Studie teilnehmen zu können, mussten die potenziellen Teilnehmenden eine Reihe von Zugangsvoraussetzungen erfüllen. Für die MRT-Scans im Rahmen der Studie wurden ein gutes Gehör, keine Klaustrophobie und Rechtshändigkeit vorausgesetzt. Ebenso wurden Stress- und Traumascreenings durchgeführt, um das Stresslevel der potenziellen Teilnehmenden einschätzen zu können. Diese Screenings beruhen auf Selbstaussagen der Teilnehmenden und es kann nicht ausgeschlossen werden, dass die Teilnehmenden hier eher ‚erwünschte' Antworten als Antworten, die tatsächlich das eigene Stresslevel reflektieren, gegeben haben. Die Teilnehmenden sollten darüber hinaus zwischen 18 und 35 Jahre alt sein und Arabisch als L1 sprechen. Dabei wurde zusätzlich eingeschränkt, dass die Teilnehmenden aus einem der arabischen Länder der

110 Als Abbrecherinnen und Abbrecher werden hier alle Teilnehmenden bezeichnet, die die Teilnahme an der Studie zu einem bestimmten Zeitpunkt vor dem Abschluss der ersten Studienphase nicht fortgesetzt haben.

111 An dieser Stelle sei den vielen Mitarbeiterinnen und Mitarbeitern des Flüchtlingsrats e. V. Leipzig ganz herzlich gedankt, die sowohl die Forschenden, aber vor allem die Geflüchteten in Leipzig mit großem Einsatz unterstützen und die geplante Studie in verschiedenen Gemeinschaftsunterkünften im Leipziger Raum im April 2016 publik machten.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Levante kommen sollten, um den Dialekthintergrund möglichst vergleichbar zu gestalten.[112] Darüber hinaus sollten die Teilnehmenden monolingual aufgewachsen bzw. ihre Sprachbiografien nicht durch frühe Bilingualität gekennzeichnet sein. Als frühe Bilingualität wurde in der vorliegenden Studie ein simultaner Spracherwerb von zwei (oder mehr) Sprachen vor dem zweiten Geburtstag definiert (vgl. Grimm 2012: 195).[113] Eine weitere Voraussetzung war ein ‚hoher' Grad an (schulischer bzw. universitärer) Vorbildung, d. h. die potenziellen Teilnehmenden sollten neben dem Arabischen ebenfalls lateinische Buchstaben lesen und schreiben können bzw. einen hohen Grad an Alphabetisierung und Literalisierung aufweisen und wenn möglich über einen Schulabschluss verfügen.[114] Dieses Kriterium wurde durch Selbstauskunft erhoben.

Zuletzt war Voraussetzung, dass die Teilnehmenden keine oder nur geringe Vorkenntnisse im Deutschen hatten. Die Gruppe der Teilnehmenden sollte auf diese Weise mit den Teilnehmenden in einem Modul 1 eines Integrationskurses vergleichbar sein. Wie bei Integrationskursen wurde dieser Umstand in der vorliegenden Studie durch Einstufungstests zu Beginn der Studie überprüft (vgl. Hünlich et al. 2018: 12). Zur Einstufung der Sprachkenntnisse der Teilnehmenden wurden die Einstufungstests des Lehrwerks ‚Menschen' und ein Modelltest ‚Hören' aus dem Goethe-Zertifikat A1 benutzt.[115] Ein zu hohes Ergebnis in diesen Einstufungstests hatte zur Folge, dass Personen nicht an der Studie teilnehmen konnten. Als Grenzwert für eine mögliche Teilnahme wurde das Niveau A1.1 angesetzt. Es muss angemerkt werden, dass nicht immer klare Entscheidungen getroffen werden konnten und im Zweifel eher eine Teilnahme der betreffenden Personen präferiert wurde. Bei der Konzentration auf Teilnehmende aus der

112 Dies ist für das Arabische sicherlich kein einfaches Unterfangen, da Arabisch einerseits eine plurizentrische Sprache ist und sich die L1-Sprecher/-innen des Arabischen andererseits in einer grundsätzlichen Diglossiesituation zwischen dem schriftsprachlichen modernen Arabisch und den gesprochenen und regionalen Varianten befinden (vgl. zur Diglossie zwischen arabischer Hochsprache und Umgangssprache z. B. Riedner / Kassem 2010; Kadamani 2012: 79–84).

113 In der Literatur existieren in diesem Zusammenhang zum Teil abweichende Definitionen. Ahrenholz (2020) bezeichnet mit dem Begriff etwa lediglich den Spracherwerb von zwei oder mehr Sprachen in den ersten Lebensjahren und bezeichnet diesen Umstand als *„Bilingualismus* oder *doppelte[n] Erstspracherwerb* (ebd.: 5; Kursivsetzungen im Original). Dem Begriff des Bilingualismus stellt Ahrenholz die Definition von Mehrsprachigkeit gegenüber, als Bezeichnung „für alle Formen multipler Sprachkompetenz" (ebd.). Zusätzlich weicht die englischsprachige Beschreibung der Kontexte zum Teil von den Beschreibungen im Kontext DaF/DaZ und der deutschsprachigen Fremdsprachenforschung ab.

114 Dieses Kriterium wurde aufgrund des Kriegs in Syrien oder der kriegsähnlichen Situation im Irak flexibel angewandt. Beispielsweise nahmen Lernende an der Studie teil, die zwar noch keinen Schulabschluss hatten, aber vor Ausbruch des Krieges kurz davorstanden.

115 Die Links zu den beiden Tests finden sich in Tabelle 3.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Levante spielte einerseits, wie angedeutet, der Wunsch eine Rolle, eine Gruppe von Teilnehmenden mit möglichst homogenem Sprachhintergrund auszuwählen, andererseits standen ethische Bedenken dahinter. Es sollte sichergestellt werden, dass die potenziellen Teilnehmenden aus Ländern mit ‚guter Bleibeperspektive' stammten oder ihnen Flüchtlingsstatus bzw. subsidiärer Schutz zuerkannt worden war.[116] Ein solcher Status ermöglichte den Teilnehmenden, sich jederzeit für einen durch das Bundesamt für Migration und Flüchtlinge (BAMF) angebotenen Sprachkurs zu entscheiden und einen solchen wahrnehmen zu können. In jedem Fall sollte ausgeschlossen werden, dass die Teilnahme an der Studie der einzige Weg zur Teilnahme an einem Sprachkurs war.[117]

Zusätzlich unterzeichneten alle Teilnehmen+den eine Einverständniserklärung darüber, dass die erhobenen persönlichen Daten in anonymisierter Form zur Auswertung der Studie herangezogen werden und veröffentlicht werden dürfen. Eine genaue Übersicht über die eingesetzten Fragebögen und Instrumente findet sich in Tabelle 3:

Zugangsvoraussetzung	erhoben durch	Ausschlusskriterium	verfügbar
Arabisch als L1	Anmeldeform für Testpersonen (MPI)	Arabisch als Zweitsprache bzw. Arabisch als Fremdsprache	MPI-intern
Keine Einschränkungen beim Gehör	Anmeldeform für Testpersonen (MPI)	Selbstaussage	MPI-intern
Keine Klaustrophobie	Anmeldeform für Testpersonen (MPI)	Selbstaussage	MPI-intern
Rechtshändigkeit	Edinburgh Handedness Inventory (Oldfield 1971; in arabischer Übersetzung)	Unklare Händigkeit oder Linkshändigkeit aufgrund des Fragebogens	MPI-intern
18–35 Jahre alt	Anmeldeform für Testpersonen (MPI)	Teilnehmende zu Beginn der Studie außerhalb dieses Altersbereichs	MPI-intern

116 Das Bundesamt für Migration und Flüchtlinge (BAMF) formuliert eine ‚gute Bleibeperspektive' für „Menschen, die aus Herkunftsländern mit einer Schutzquote von über 50 Prozent kommen" (BAMF 2021), unter: https://www.bamf.de/SharedDocs/FAQ/DE/IntegrationskurseAsylbewerber/001-bleibeperspektive.html?nn=282388 [23.02.2023]. Im Jahr 2016 traf dies neben Menschen aus Syrien auf Menschen aus Eritrea, Irak, Iran und Somalia zu (vgl. BAMF / EMN et al. 2016: 54).

117 Da viele Kursanbieter in den Jahren 2015 und 2016 erst neue Kurskapazitäten schaffen mussten, konnte zum Teil nicht allen Geflüchteten sofort ein Platz in einem Integrationskurs zur Verfügung gestellt werden, was den Druck auf die Geflüchteten auf der Suche nach Sprachkursangeboten zusätzlich erhöhte.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Zugangsvoraussetzung	erhoben durch	Ausschlusskriterium	verfügbar
Keine frühe Bilingualität	Anmeldeform für Testpersonen (MPI)	Selbstaussage; definiert als kein simultaner Spracherwerb beider Sprachen vor dem zweiten Geburtstag (Grimm 2012: 195)	MPI-intern
Keine / Geringe Vorkenntnisse der deutschen Sprache	Einstufungstest	Einstufungstest aus *Menschen* mit einem Ergebnis nicht höher als A1.2; Hörteil des Goethe-Zertifikat A1 Start Deutsch 1 (Goethe-Institut) mit einem Ergebnis nicht höher als 13/15 Punkten)	Einstufungstest Menschen A1.1 und A1.2 unter: https://hueber.de/media/36/538600_Einstufungstest_A1_Teil_01.pdf [23.02.2023]; Goethe-Zertifikat Start Deutsch 1 Modellsatz unter: https://www.goethe.de/de/spr/kup/prf/prf/sd1/ueb.html [23.02.2023].
‚Hoher' Grad an Vorbildung	Anmeldeform für Testpersonen (MPI)	Selbstaussage: Schule beendet oder in Abschlussklasse bei Abbruch des Schulbesuchs; Begonnenes oder abgeschlossenes Studium	MPI-intern
Stresslevel	Trauma Screening Questionnaire (TSQ); PCL–5 (PTSD Screening)	Überschreitung von Grenzwerten	TSQ unter: https://www.ptsd.va.gov/professional/assessment/screens/tsq.asp#obtain [23.02.2023]; PCL–5 unter: https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp [23.02.2023].

Tabelle 3: Zugangsvoraussetzungen und eingesetzte Testverfahren zu Beginn der Studie

3.1.2 Teilnehmende innerhalb der Gesamtstudie

Die Gruppe der Teilnehmenden war darüber hinaus, wie oben kurz im Kontext der ‚guten Bleibeperspektive' angedeutet, durch bestimmte gemeinsame Merkmale gekennzeichnet. Alle Teilnehmenden waren erwachsene Personen mit

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Flüchtlingsstatus nach § 3 Abs. 1 AsylG oder Personen, denen nach §4 Abs. 1 AsylG subsidiärer Schutz zuerkannt wurde.[118] Sie stammten aus arabischsprachigen Ländern der Levante, d.h. aus Syrien, Jordanien und dem Libanon. Zusätzlich gab es Teilnehmende aus dem Irak.

In dieser Studie wurde auf eine erstsprachliche Kontroll- bzw. Vergleichsgruppe verzichtet.[119] Einerseits ergab sich dies aus der theoretischen Perspektive, die dieser Arbeit zugrunde liegt. Zweitsprachliche Entwicklung wird als komplexdynamisches System gesehen, welches nicht teleologisch zu betrachten ist, sondern in seiner graduellen und longitudinalen Entwicklung analysiert werden sollte (vgl. zu dieser Perspektive Kapitel 2.5). Die teleologische Orientierung auf eine L1-Norm hin, deren Erreichen dann im Grunde ausschließlich über bestimmte Korrektheitsquoten attestiert werden kann (vgl. Ortega 2014: 197–198), beinhaltet stets die Gefahr der „comparative fallacy" (Bley-Vroman 1983: 14–16) zu unterliegen. Der Verzicht auf eine L1-Kontrollgruppe in Bezug auf die Sprachdaten basierte zusätzlich auf der Annahme, dass die zu beobachtende Variabilität innerhalb der Daten von L1-Sprechenden in Bezug auf die eher basalen Wortstellungsregeln, die in dieser Studie analysiert werden, zu gering ausfallen könnte, um sinnvolle Vergleiche anzustellen. Vielmehr scheint es so zu sein, dass L1-Schreibende die hier relevanten Wortstellungsregeln spätestens mit Abschluss der Grundschule beherrschen (vgl. zu diesem Punkt Kapitel 2.1; vgl. aber einschränkend Ulrich (2017)).

Da alle Teilnehmenden aus ihren Heimatländern geflüchtet sind, fanden sie sich plötzlich in einer Situation, in der ihre bisherigen Lebensleistungen im Ankunftsland Deutschland kaum noch eine Rolle spielten und familiäre und freundschaftliche Beziehungen neu aufgebaut werden mussten. Hinzu kam ein Kontrollverlust über „zentrale Bereiche ihres Lebens" (Drumm / Henning 2016: 11). Dies war besonders evident bei der Zuteilung in ein Aufnahmelager und der Aufnahme des Asylverfahrens (vgl. ebd.) und wird insbesondere von Geflüchteten mit ‚hohem' Bildungsgrad sehr einschneidend wahrgenommen (vgl. Coenen 2017: 75). Darüber hinaus wurde dies alles in einer Situation „migrationsbedingter lebensweltlicher Mehrsprachigkeit" (vgl. Gogolin 1988; vgl. Lüdi / Py 2009; vgl. Geist 2021) erfahren (zu einer Diskussion der Unterschiede zwischen migrationsbedingter und ‚europäischer' Mehrsprachigkeit vgl. etwa Altmayer 2022: 22). Migrationsbedingte Mehrsprachigkeit wird vor allem in Situationen erlebt, in denen Migra-

118 Das Bundesamt für Migration und Flüchtlinge informiert auf seinen Seiten sehr ausführlich über den Ablauf der verschiedenen Verfahren im Kontext einer Migration oder Flucht nach Deutschland und über die verschiedenen Schutzformen, die im Rahmen dieser Verfahren gewährt werden können. Weitere Informationen finden sich unter:
https://www.bamf.de/DE/Themen/AsylFluechtlingsschutz/AblaufAsylverfahrens/Schutzformen/schutzformen-node.html [23.02.2023].

119 Für die psychologischen Tests und für die MRT- und fMRT-Daten wurden L1-Kontrollgruppendaten erhoben.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

tionsbewegungen vielfältige Sprachkontakte zur Folge haben, die sich auf individueller und gesellschaftlicher Ebene niederschlagen (vgl. Geist 2021: 79).

Die Teilnehmenden brachten unterschiedliche Flucht- und Verlusterfahrungen bzw. Migrationserfahrungen in den Unterricht mit. Hier ist durchaus davon auszugehen, dass diese Erfahrungen in vielen Fällen traumatischer Art waren und dadurch Einfluss auf das Lernen einer neuen Sprache haben konnten. Mögliche psychologische Faktoren im Hintergrund könnten etwa Kriegs- und Verlusterfahrungen, Gewalt, Folter, lange und schwierige Fluchtwege, Stress durch Bürokratie und Asylbestimmungen im Ankunftsland, sprachliche Anforderungen, gesetzliche Bestimmungen, Unterbringung in Gemeinschaftsunterkünften, fehlende Privatsphäre, fehlende Unterstützung bzw. Betreuung, ein hohes Maß an gefühlter und tatsächlicher Fremdbestimmung, der erschwerte Zugang zu medizinischer Versorgung und Angst um im Kriegsland zurückgebliebene Angehörige gewesen sein (vgl. Schwald 2017: 28–29). Darüber hinaus müssen beim Lernen einer Zweitsprache Lernstrategien, die in Bildungskontexten im Heimatland angeeignet wurden, oft nach langer „Nicht-Inanspruchnahme" (Schwald 2017: 31) wieder reaktiviert werden, was sehr große geistige Kapazitäten voraussetzt. Die Gruppe der Teilnehmenden gehörte folglich zu einer Gruppe mit erhöhtem Risikofaktor für PTSD (*Post traumatic stress disorder*) (auf Deutsch PTBS) oder anderen durch Traumata bzw. psychische Stresssituationen begünstigte Krankheitsbilder, die gleichzeitig in dieser Gruppe oft seltener erfasst bzw. diagnostiziert werden (vgl. Belz et al. 2017: 596). Zu den Begleiterscheinungen posttraumatischer Belastungsstörungen gehören neben Schlafstörungen zum Beispiel Belastungsreaktionen wie depressive oder dissoziative Zustände (vgl. u.a. Perry 2006; Schwald 2017; Coenen 2017; Belz et al. 2017; Özkan / Althoff 2019). Es wurden in der vorliegenden Studie keine Daten dazu erhoben, ob bei den Teilnehmenden PTSD oder andere psychische Erkrankungen diagnostiziert wurden. Um die Teilnehmenden der vorliegenden Studie so weit wie möglich zu entlasten bzw. nicht zusätzlich zu belasten, wurden nur in sehr beschränktem Umfang Daten zur Fluchtgeschichte der Teilnehmenden oder zur Einreise nach Deutschland bzw. in die Europäische Union erhoben. Um die Persönlichkeitsrechte der Teilnehmenden zu schützen, wurden einige dieser Daten außerdem nicht ausgewertet und nicht veröffentlicht. Dies ist sicher eine methodische Einschränkung der vorliegenden Studie. Es sollten jedoch nicht die Auswirkungen psychischer Erkrankungen oder Traumata auf den Spracherwerb untersucht werden. Die Gefahr einer Retraumatisierung durch ein erzwungenes Erinnern an traumatische Ereignisse sollte vermieden werden (vgl. Coenen 2017: 77f.). Hier kann die vorliegende Studie aus den genannten Gründen leider nur wenig zur weiteren Erforschung dieses Kontextes beitragen (vgl. hierzu Kapitel 3.1.6 zu Limitationen und Begrenzungen der Studie).[120]

120 Es wurde versucht, die Teilnehmenden so eng wie möglich und wie von diesen gewünscht zu begleiten und Kontakte zu betreffenden Stellen und Gesprächspartnerinnen und Gesprächspartnern herzustellen.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Insgesamt scheint die Erforschung von Lernprozessen vor dem Hintergrund schwerwiegender Traumatisierungen in vielen Studien zum Zweitspracherwerb bzw. zur Zweitsprachentwicklung oft nur am Rande beachtet zu werden. Dieser Themenkomplex stellt daher ein großes Desiderat dar. Es gibt nur wenige Studien, die diesen Kontext genauer beleuchten. Johnsen / Asbjørnsen (2009) berichten in einer Studie im norwegischen Kontext, dass die Leistungsfähigkeit beim Lernen von Wortlisten bei Teilnehmenden mit PTSD zwar geringer sei als bei Teilnehmenden ohne PTSD, sich die Lernkurven in beiden Gruppen aber stark ähnelten. Søndergaard (2017) stellt eine Studie aus dem schwedischen Kontext vor, die zu ähnlichen Ergebnissen kommt und fordert, dass der Sprachunterricht mit Gruppen, in denen Lernende mit PTSD lernen, bestimmte Voraussetzungen erfüllen müsse. Lehrende sollten dafür sensibilisiert werden, dass sie mit Lernenden arbeiten, die unter Umständen große Gedächtnisschwierigkeiten haben. Im Unterricht sollten viele Wiederholungsphasen in einem strukturierten Spiralcurriculum geplant werden (vgl. Søndergaard 2017: 18). Coenen (2017) hat für den DaZ-Kontext eine qualitative Interviewstudie zu Traumafolgen durchgeführt. In Rahmen dieser Erhebung interviewt sie sechs Expertinnen und Experten, um Perspektiven der Traumatherapie für den DaZ-Unterricht auszuloten. Für den Bereich DaZ konstatiert sie insgesamt ein undifferenziertes Bild von Traumata, welches sich etwa in verschiedenen Handreichungen für den Kontext DaZ niederschlägt, die ein sehr enges Traumabild anlegen (vgl. Coenen 2017: 65). Davon ausgehend fordert sie eine differenziertere Perspektive auf die potenziell sehr verschiedenen Erscheinungsformen von Traumata für den Kontext DaZ und den Umgang mit traumatisierten Lernenden, einen „traumatic turn" (ebd.: 95). Coenen (2017) macht darauf aufmerksam, dass Therapieangebote in Deutschland oft ausschließlich auf Deutsch verfügbar seien, was Menschen, die ohne Sprachkenntnisse in Deutschland ankommen, von der Inanspruchnahme von solchen Angeboten zu einem Zeitpunkt ausschließe, an dem diese essenziell wären. Die Lebenssituation von Geflüchteten kann darüber hinaus oft „durch Mehrfachbelastung gekennzeichnet" (Schwald 2017: 29) sein. Unterrichts- und Lernsituationen könnten verstärkten Stress verursachen; einmal aufgrund traumatischer Erfahrungen und dann wegen möglicherweise ‚stressigen' Erfahrungen in der eigenen Schulzeit (vgl. Perry 2006: 22). Mehr noch könnten körperliche und psychische Beschwerden als Folge von Traumata schlimmer werden, wenn der Stresspegel der Lernenden steigt (Coenen 2017: 68). In unterrichtlichen Lernkontexten „können es harmlose Situationen (eine Frage, ein Flugzeuggeräusch, ein Blick…) sein, die betroffene Personen in Angst und Schrecken versetzen" (Schwald 2017: 30). Das bedeutet gleichzeitig, dass im Unterricht potenziell Retraumatisierungen ausgelöst werden können und dass „weder die traumatisierte Person noch die Lehrperson in der Lage sind, diese Trigger zu kontrollieren" (Coenen 2017: 77). Perry (2006) macht darauf aufmerksam, dass Lernende sich in einem dauerhaften, niederschwelligen Angstzustand befänden und daher Informationen nicht immer aufnehmen könnten (vgl. Perry 2006: 25). Falls jedoch ein gelingender Unterricht angeboten werden

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

könnte, hätte dieser unter Umständen eine stabilisierende Funktion (vgl. Coenen 2017: 75).

Bei der Gruppe der Teilnehmenden handelte es sich einerseits um Lernende, die potenziell Integrationskurse besucht hätten (vgl. zu den gesetzlichen Rahmenbedingungen der Integrationskurse Gamper / Schroeder 2021: 69f.).[121] Andererseits handelte es sich um Lernende, die aufgrund der Vorauswahl als literal erfahrene Lernende beschrieben werden können (vgl. Marx 2017; vgl. Kapitel 3.1.1 zur Vorauswahl). Daher sind die Daten, die 2016–2017 vom Leibniz-Institut für Deutsche Sprache (IDS) und dem Goethe-Institut im Rahmen des Projekts „Deutsch im Beruf: Die sprachlich-kommunikative Integration von Flüchtlingen" erhoben wurden, von besonderer Relevanz, um die Teilnehmenden der vorliegenden Studie besser einzuschätzen. Hünlich et al. (2018) vergleichen in ihrer Studie die sprachlich-sozialen Hintergründe von Geflüchteten im Kontrast zu anderen Zugewanderten (vgl. Hünlich et al. 2018: 9) mithilfe von Clusteranalysen.[122] Cindark / Hünlich (2019) geben im Rahmen derselben Datenerhebung als wesentliches Ergebnis der Studie die „enorm heterogene Zusammensetzung der Integrationskurse im Hinblick auf Herkunftsregionen, Sprachhintergründe, Alter, Aufenthaltsdauer sowie Bildungs- und Berufsbiografien" (Cindark / Hünlich 2019: 19) an.

Durch die relativ klar geregelten Zugangsvoraussetzungen der MPI CBS/HI-Studie ist im Gegensatz zur Studie des IDS und des Goethe-Instituts zunächst von einer vergleichbareren bzw. ‚homogeneren' Gruppe von Teilnehmenden auszu-

121 Gamper / Schroeder (2021) geben eine Zusammenfassung über Lehr- und Lernkontexte und die generelle Situation für erwachsene DaZ-Lernende in Deutschland, die sie mit einem historischen Überblick über die Entwicklungen in diesem Bereich ergänzen.

122 Cindark / Hünlich (2019) können aus den Daten von mehreren hundert Lernenden in Integrationskursen mithilfe von Clusteranalysen fünf verschiedene Gruppen von Teilnehmenden rekonstruieren (vgl. Cindark / Hünlich 2019: 20f.): Die „Spätausgewanderten", die „Unterpriviligierten", „die Hoffnungsträger/innen", die „Durchstarter/innen" und die „Langmigrierten". Die Autoren kommen anhand ihrer Daten zu dem Ergebnis, dass vor allem „Bildung, Alter sowie die Anzahl der [bereits gelernten] Fremdsprachen einen hochsignifikanten Einfluss auf die Wahrscheinlichkeit haben, dass eine teilnehmende Person A2 oder besser erreicht" (Cindark / Hünlich 2019: 24). In ihren Daten können die Forschenden keine Effekte für die Arbeitserfahrung vor der Ankunft in Deutschland, frühe Mehrsprachigkeit oder sprachliche Verwandtschaft der Familiensprache mit dem Deutschen nachweisen (vgl. ebd.). Ebenso kann keine Wirkung des Aspekts Flucht als Migrationsmotiv nachgewiesen werden (ebd.). In der untersuchten Lernendengruppe erreichen nur sehr wenige Lernende nach Abschluss der sechs Module des Integrationskurses in ihren mündlichen Kompetenzen das angestrebte Niveau B1 (2 %) (vgl. ebd.: 23). Ebenso erreichen die meisten der Teilnehmenden das Niveau A2 nicht, sondern wurden von den Bewertenden auf ein Niveau unter A2 eingestuft (55,5 %) (vgl. ebd.: 23).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

gehen. Die hier vorliegende MPI CBS/HI-Studie war ursprünglich auf 90 Teilnehmende ausgelegt. Letztlich starteten 94 Teilnehmende die Studie, die auf insgesamt sechs Sprachkurse verteilt waren.[123] Diese 94 Teilnehmenden waren im Schnitt zu Beginn des Unterrichts 24,42 Jahre alt (Median (*Mdn*) = 24; Mode = 20; Standardabweichung (SD) = 4,67; n=94).[124] Da geringe bzw. keine Vorkenntnisse im Deutschen eine Zugangsvoraussetzung bildeten, fällt das Alter zu Beginn der Studie in diesem Fall für die Teilnehmenden mit der Variable *Age of Onset* (AoO) bzw. dem Startpunkt des Sprachkontakts zusammen.[125] 74 der 94 Teilnehmenden gaben ihr Geschlecht als männlich an, 20 als weiblich. Als weitere Sprachen (L2) wurden von den Teilnehmenden genannt: Englisch (n=45), Kurdisch (n=2) und Spanisch (n=1). 16 Teilnehmende führten keine Kenntnisse einer weiteren Sprache an.[126] Außerdem wurden von Teilnehmenden, die L2-

123 Für die Fragestellungen der vorliegenden Arbeit wurden jeweils Teilstichproben aus der gesamten Stichprobe der Teilnehmenden gebildet. Erklärungen dazu, wie diese Stichproben gebildet wurden, finden sich in Kapitel 3.2.1 und in Kapitel 3.3.1. Außerdem wird im Kapitel zu den Limitationen der Studie die Gruppe der Abbrecher/-innen ins Verhältnis zu den einzelnen Teilstichproben gesetzt und genauer analysiert (vgl. Kapitel 3.1.6).

124 Zusätzlich zum arithmetischen Mittel (bzw. zum Mittelwert) werden in dieser Arbeit in der Regel weitere Lageparameter wie der Median, der Modalwert (*Mode*) und Streuungsparameter wie die Standardabweichung angegeben, um die Verteilung der beobachteten Daten besser einschätzen zu können.
Der **Median** bezeichnet den mittleren Wert eines nach Größe geordneten Satzes von Messwerten. Die Hälfte der Messwerte liegt unterhalb des Medians, die Hälfte der Messwerte liegt oberhalb des Medians. Der Median stellt damit gleichzeitig das ½-Quantil dar. Bei einer geraden Anzahl von Messwerten innerhalb eines Datensatzes ist der Median der Durchschnitt der beiden Werte, die jeweils um den angenommenen mittleren Wert liegen (vgl. Field 2018: 1025; Brezina 2018: 10f.). Der **Modalwert** bzw. Mode bezeichnet den am häufigsten auftretenden Wert innerhalb eines Datensatzes (vgl. Field 2018: 1025). Die **Standardabweichung** „ist eine Schätzung der durchschnittlichen Variabilität (Streuung) eines Datensatzes, die in denselben Messeinheiten wie die Originaldaten gemessen wird. Sie ist die Quadratwurzel aus der Varianz" (Field 2018: 1037; [Übersetzung MS]). Die **Varianz** ist ebenfalls ein Streuungsparameter, das auf dem arithmetischen Mittel beruht und „gibt an, wie die Werte durchschnittlich um das arithmetische Mittel streuen" (Gültekin-Karakoç / Feldmeier 2014: 190). Sie ist „der Mittelwert aus der Variation, d.h. aus den quadrierten Abweichungen aller Messwerte vom arithmetischen Mittel" (ebd.).

125 Dies gilt ebenso für die Altersangaben der Subgruppen, die gebildet wurden, um die Forschungsfragen zu beantworten (vgl. Kapitel 3.2.1 für die Gruppendaten und Kapitel 3.3.1 für die Einzelfallanalysen).

126 Viele der Teilnehmenden der Studie gaben zwar Englischkenntnisse an, erreichten aber gleichzeitig nur sehr geringe Ergebnisse in einem durchgeführten Englischtest. Gleichzeitig gaben sie die Lernzeit und die Verwendung des Englischen im persönlichen Sprachgebrauch als sehr gering an. Von einer Sprachenfolge L1 Arabisch → L2 Englisch → L3 Deutsch kann daher nicht grundsätzlich aus-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Kenntnisse angegeben haben, folgende zusätzliche Fremdsprachenkenntnisse (L2) berichtet: Französisch (n=8), Türkisch (n=7), Italienisch (n=1) und Spanisch (1=n). Für 30 Teilnehmende liegen keine Daten zur Sprachbiografie vor, da der Fragebogen zum Sprachhintergrund nicht direkt zu Beginn der Studie erhoben wurde und viele der Teilnehmenden die Teilnahme bereits wieder abgebrochen hatten (vgl. das Kapitel 3.1.6.3).

Hünlich et al. (2018) berichten in ihrer Studie ein Durchschnittsalter geflüchteter Lernender in Integrationskursen von 30,1 (vgl. ebd.: 11). Damit sind die vergleichbaren Lernenden in Integrationskursen ein wenig älter als in der vorliegenden Studie. Die Geschlechterverteilung hingegen ist fast identisch. Hünlich et al. (2018) ermitteln hier eine Verteilung von 80 % männlichen und 20 % weiblichen Lernenden. Weiterhin konstatieren die Forschenden, dass die Teilnehmenden 10–24 Monate in Deutschland seien, bevor sie den Integrationskurs besuchen (vgl. ebd.: 12) und dass in den Integrationskursen der Grad an Vorbildung relativ hoch sei, da vorab „eigens Kurse für Analphabeten und Schriftlerner/innen angeboten" (ebd.: 13) würden. Die Teilnehmenden der Studie von Hünlich et al. verbrachten im Schnitt zwischen 10 und 16 Jahre in Ausbildungskontexten. Nur ein Viertel der teilnehmenden „Geflüchteten und ein Sechstel der anderen Zugewanderten verbrachten weniger als zehn Jahre in der Schule" (ebd.: 13). In der IDS-Goethe-Studie sind Lernende mit Arabisch als L1 die zahlenmäßig größte Gruppe von Lernenden. Insgesamt geben dort 56 % der Lernenden Arabisch als L1 an (vgl. ebd.: 19). Dabei stammen 71 % der Befragten mit Arabisch als L1 aus Syrien (vgl. ebd.). Als weitere L2 wird hier ebenfalls am häufigsten Englisch genannt (vgl. ebd.: 24). Dies rückt die Gruppe arabischsprachiger Deutschlernender, die ebenfalls in der vorliegenden Gruppe im Fokus der Erhebung standen, deutlich in den Fokus.

3.1.3 Longitudinaler Aufbau und zeitlicher Ablauf der Studie

In diesem Kapitel werden der longitudinale Aufbau und der daraus resultierende zeitliche Ablauf der Studie beschrieben.[127] In einer longitudinalen Studie werden mehrfach Daten derselben Teilnehmenden erhoben, um Veränderungen im Verhalten dieser Teilnehmenden sichtbar und nachvollziehbar machen zu können. Solche Daten werden als Zeitreihen oder Zeitreihendaten beschrieben (vgl. zur Zeitreihenanalyse Backhaus et al. 2018: 126–161). Eine Zeitreihe ist definiert als „eine Menge von Werten einer Variablen Y […], die im Zeitablauf gemessen wurden und gemäß der Zeit geordnet sind" (ebd.: 127). Longitudinale Studien haben den Vorteil, dass sie Beziehungen zwischen verschiedenen Variablen

gegangen werden. Einige der Teilnehmenden berichteten ein hohes Niveau in anderen Sprachen wie Türkisch oder Kurdisch, die nicht einzeln getestet werden konnten.

127 Siehe hierzu den Zeitplan der Gesamtstudie im Anhang.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

präziser erfassen können als Querschnittsstudien, weil sie Informationen über den zeitlichen Entwicklungsverlauf der beobachteten Variablen für dieselben beobachteten Subjekte liefern können (vgl. Barkaoui 2014: 66). Darüber hinaus dienen Zeitreihendaten der Prognose von Variablen, das heißt „der Schätzung von Werten dieser Variablen für zukünftige Zeitpunkte oder Perioden" (Backhaus et al. 2018: 127). Im Gegensatz zu Querschnittsstudien, die oft lediglich Gruppendurchschnitte aggregieren können, um Aussagen zu einem bestimmten Zeitpunkt innerhalb der angenommenen Entwicklung von Individuen zu machen, ermöglichen longitudinale Studien die Untersuchung intraindividueller Veränderungen (vgl. Collins 2006) und die Beobachtung intraindividueller Variabilität in der tatsächlichen Entwicklung. Longitudinale Studien sind daher geeignet, Entwicklungsverläufe nachzuzeichnen (vgl. Schramm 2018: 36).

Längsschnittdaten sind nach Ortega / Iberri Shea 2005 durch vier grundsätzliche Aspekte gekennzeichnet: Eine bestimmte Studienlänge, die Sammlung von Daten in mehreren Wellen (*multiwave-data-collection*), die planmäßige Untersuchung von Veränderung (*Change by Design*) und die Herstellung von Beziehungen zwischen den beobachteten Phänomenen.[128] Grundsätzlich stehen longitudinale Studien vor einer Reihe von notwendigen Entscheidungen, die das Studiendesign betreffen. Barkaoui (2014) nennt die Anzahl und Auswahl der Teilnehmenden, die Länge der Studie, die Definition von untersuchter Zeit, die zu messenden Variablen und zuletzt die zeitliche Anordnung und die Anzahl von Messpunkten (vgl. Barkaoui 2014: 68; vgl. außerdem das nächste Unterkapitel 3.1.4 zum Materialzugang und zu den Messpunkten). Zeitliche Fragestellungen sollten vor Studienbeginn und hinsichtlich einer zugrundeliegenden Theorie ausgewählt werden (vgl. Barkaoui 2014: 68). Für die vorliegende Studie bedeutet dies einige methodische Herausforderungen. Einerseits müssen Messabstände so gewählt werden, dass sie aus neurowissenschaftlicher Sicht hinsichtlich der MRT und fMRT-Scans Sinn ergeben, andererseits müssen sie sich an fremdsprachendidaktischer Progression über verschiedene GER-Niveaus und einem angemessenen Curriculum orientieren. Zusätzlich müssen Ferienphasen eingeplant und im Hinblick auf die Gruppe an Teilnehmenden, die mehrheitlich muslimischen Glaubens ist, ein Unterrichtsplan für den Unterricht während des muslimischen Fastenmonats Ramadan geplant werden.

Erste Studienphase

Die vorliegende Studie wurde von Mai 2016 bis September 2017 durchgeführt und war in zwei große Studienphasen geteilt. Die erste Studienphase wurde zwischen Mai 2016 und Februar 2017 durchgeführt, die zweite Phase von März 2017 bis September 2017. In beiden Phasen wurden an fünf Tagen pro Woche

128 Barkaoui (2014) weist darauf hin, dass es nur für zwei der Aspekte einen Konsens in der Forschungsliteratur zu geben scheint, nämlich für die Sammlung von Daten in verschiedenen Wellen und die Untersuchung von Veränderung (vgl. ebd.: 66).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

jeweils fünf Unterrichtsstunden täglich für die Lernenden angeboten. Die Teilnehmenden waren in Studienphase 1 in sechs verschiedene Lerngruppen (LG) mit insgesamt 94 Teilnehmenden aufgeteilt.[129] In jeder Lerngruppe waren 12–19 Teilnehmende (LG 1: 17; LG 2: 13; LG 3: 16; LG 4: 15; LG 5: 19; LG 6: 14).[130] Drei der Gruppen lernten am Vormittag von 9:00–13:30 Uhr und drei der Gruppen am Nachmittag von 14:00–18:30 Uhr. In jedem der Kurse unterrichteten jeweils drei Lehrkräfte, um zu garantieren, dass die Lernenden unterschiedlichen sprachlichen Input erhielten.[131] Das kurstragende Lehrwerk war für die Niveaustufen A1, A2 und B1 „Menschen" aus dem Hueber Verlag.[132] In der zweiten Studienphase wurde für die höheren Niveaustufen B1+, B2 und C1 das Lehrwerk „Sicher" eingesetzt.[133] Es wurden jeweils die Kurs- und die Arbeitsbücher eingesetzt. In den Kursen wurde ein kommunikativer Ansatz verfolgt. Dies bedeutete vor allem, dass die Kursinhalte sich an den kommunikativen Bedürfnissen der Lernenden orientieren sollten. Grammatische Inhalte wurden weitgehend in konkrete kommunikative Kontexte eingebunden und nicht isoliert vermittelt. Eine Ausnahme bildeten hier zum Teil Inhalte, die sich auf die experimentelle Manipulation innerhalb der Studie bezogen (vgl. zur experimentellen Manipulation Kapitel 3.1.4). Für alle Lerngruppen wurden die zu vermittelnden Kursinhalte in jeder Woche durch den Studienkoordinator vorbereitet und den Lehrenden in Form eines Rahmenplans zur Verfügung gestellt. In diesem Rahmenplan wurden die einzelnen Übungen und Materialien sowie die zu bearbeitenden Themenkomplexe aus dem Lehrwerk festgelegt, die in der jeweiligen Woche unterrichtet werden sollten. Die Lehrenden planten ihren Unterricht dann ausgehend von diesen Rahmenplänen. Dabei waren sie in der konkreten Ausgestaltung des Unterrichts relativ frei. Von diesen sechs Lerngruppen hatten im Rahmen der der Studie zugrundeliegenden experimentellen Manipulation drei einen Schwerpunkt auf der Vermittlung von Syntax und drei einen semantischen Schwerpunkt (vgl. hierzu das Kapitel 3.1.5 zur experimen-

129 Zwei Teilnehmende absolvierten zwar alle Tests vor dem Kurs und nahmen darüber hinaus am ersten Scan teil, signalisierten dann aber direkt, dass sie nicht an der weiteren Studie teilnehmen würden. 29 Teilnehmende brachen die Studie in den ersten Wochen bzw. Monaten ab (vgl. hierzu Kapitel 3.1.6.3).

130 Die Begriffe Kurs und Lerngruppe werden im Folgenden synonym verwendet.

131 Ob die Tageszeit, zu der gelernt wurde, einen Effekt auf das Lernen hat, wurde in der vorliegenden Studie nicht untersucht. Ebenfalls wurde nicht untersucht, ob es zwischen den Lerngruppen Effekte bezüglich der in den Gruppen unterrichtenden Lehrkräfte gibt.

132 Kursbücher: A1: Evans et al. (2012); A2: Habersack et al. (2013); B1: Braun-Podeschwa et al. (2015); Arbeitsbücher: A1: Glas-Peters et al. (2012); A2: Breitsameter et al. (2013); B1: Breitsameter et al. (2015).

133 Kursbücher: B1+: Perlmann-Balme / Schwalb (2012); B2: Perlmann-Balme / Schwab (2014); C1: Perlmann-Balme et al. (2016); Arbeitsbücher: B1+: Orth-Chambah et al. (2012); B2: Perlmann-Balme et al. (2014); C1: Perlmann-Balme et al. (2016).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

tellen Manipulation). In der ersten Phase sollte das B1-Niveau nach dem GER erreicht werden. Auf dem Weg dorthin wurden zusätzlich ein A1-Test, ein Test zwischen A1 und A2 und ein A2-Test durchgeführt. Für jedes Sprachniveau waren zunächst etwa 200 Stunden Unterricht bzw. zwei Monate eingeplant. Die Planung verschob sich zum Teil, weil etwa in den Monaten des Ramadans 2016 (6.6.–5.7.2016) und 2017 (26.5.–24.6.2017) Unterricht mit reduzierter Stundenzahl durchgeführt wurde. Ziel bei der Planung war es, für jedes Sprachniveau etwa 200 Stunden Unterricht anzubieten.

In der Darstellung des Studiendesign ist zusätzlich zu beachten, dass nicht alle Teilnehmenden Phase 1 der Studie gleichzeitig begonnen haben, sondern, dass die sechs Unterrichtskurse gestaffelt nacheinander gestartet wurden (vgl. Tabelle 4).

	Mai 2016	Juni 2016	Juli 2016	Aug. 2016	Sept. 2016	Okt. 2016	Nov. 2016	Dez. 2016	Jan. 2017	Feb. 2017
Kurs 1 (Syntax, A1.1) (Start: 2.5.2016)							X			
Kurs 2 (Syntax, A1.2) (Start: 9.5.2016)							X			
Kurs 3 (Semantik; A1.1) (Start: 23.5.2016)								X		
Kurs 4 (Semantik, A1.1) (Start: 30.5.2016)								X		
Kurs 5 (Syntax, A1.1) (Start: 20.7.2016)									X	
Kurs 6 (Semantik, A1.2) (Start: 10.8.2016)										X

Tabelle 4: Übersicht über den zeitlichen Aufbau der Studie und die sechs Kursgruppen. Die farbig unterlegten Spalten zeigen den Kursverlauf. Mit einem ‚X' ist jeweils der Zeitpunkt der B1-Prüfung in den einzelnen Kursen markiert.

Phase 1 schlossen am Ende 65 Lernende ab. 62 dieser Teilnehmenden legten darüber hinaus das Goethe-Zertifikat B1 des Goethe-Instituts ab. Der gestaffelte Beginn der verschiedenen Lerngruppen führte dazu, dass die B1-Abschlussprüfungen nicht zu einem einzelnen Zeitpunkt stattfanden. Während vier der Kurse bis Dezember 2016 die B1-Prüfung ablegten, absolvierten die beiden letzten Kurse diese Prüfung im Februar 2017. Der Zeitpunkt der B1-Prüfung ist mit einem fett gedruckten X in Tabelle 4 markiert**.** Damit zu Beginn der zweiten

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Studienphase sichergestellt werden konnte, dass alle Teilnehmenden etwa auf einem ähnlichen Niveau den zweiten Studienteil starteten und damit Kursinhalte präsent gehalten werden konnten, wurde für die Teilnehmenden aus den Kursen 1–4 in der Zeit zwischen Januar und Februar 2017 in reduzierter Stundenzahl weiter Unterricht gegeben.

Zweite Studienphase

An die erste Studienphase und die kurze Zwischenphase schloss sich eine zweite Phase von März 2017 bis September 2017 an, in der der weitere Spracherwerb der Teilnehmenden untersucht wurde. In dieser Phase wurden in zwei Lerngruppen insgesamt 42 Studierende in zwei Studiengruppen für weitere 6,5 Monate unterrichtet. Voraussetzungen zur weiteren Teilnahme an der Studie waren die bestandene Prüfung zum Goethe-Zertifikat B1 und eine abgeschlossene Teilnahme an allen Scans und psychologischen Testungen bis zu diesem Zeitpunkt. Das Sprachniveau B1 nach dem GER bildete das Ausgangsniveau für den zweiten Studienteil. Ziel der zweiten Studienphase war es, das C1-Niveau nach dem GER und damit die zur Hochschulzugangsberechtigung notwendigen sprachlichen Kenntnisse zu erreichen. Die Teilnehmenden schlossen diese Studienphase entweder mit der telc B2 Prüfung oder dem TestDaF ab.

Mit Abschluss der ersten Studienphase und Beginn der zweiten Studienphase wurde zusätzlich eine umfangreiche Testbatterie aus standardisierten ACTFL-Tests eingesetzt, die zum Teil vom Institut für Testforschung und Testentwicklung e.V. Leipzig (ITT) für Deutsch adaptiert wurden.[134] Außerdem wurden verschiedene Tests eingesetzt, die direkt durch das ITT entwickelt wurden. Neben dem *ACTFL Oral Proficiency Interview computer* (OPIc) wurden der skalierte *Reading Proficiency Test* (RPT), der Lesekompetenzen testet, (vgl. Tschirner 2020; Tschirner 2015; Tschirner et al. 2015) und der *Listening Proficiency Test* (LPT), der Hörverstehen testet (Tschirner 2020a; Tschirner 2015a; Tschirner et al. 2015), durchgeführt. Beide Tests sind einerseits an die ACTFL-Proficiency Skalen angelegt und andererseits durch das ITT auf ihre Entsprechungen hinsichtlich der GER-Skalen evaluiert. Das ebenfalls durchgeführte ACTFL-OPI ist eine seit den frühen 1980er-Jahren des letzten Jahrtausends durchgeführte holistische Sprachprüfung aus dem US-amerikanischen Kontext, die mündliche Sprachproduktion testet. Das computerbasierte ACTFL-OPIc stellt eine ressourcenschonende und überall auf der Welt verfügbare Weiterentwicklung des traditionellen face-to-face OPI dar (vgl. Isbell / Winke 2019). Weiterhin wurden die durch das ITT entwickelte *Elicited Imitation Tasks* (EIT) zur Feststellung bestimmter grammatischer Kompetenzen sowie rezeptive und produktive Wort-

134 Das ITT in Leipzig wurde 2012 gegründet und beschäftigt sich mit der wissenschaftlichen Erforschung von Sprachtests und der Entwicklung, Analyse und Optimierung von Sprachtests. Dabei werden kommerzielle Testanbieter wissenschaftlich begleitet. Weitere Informationen finden sich unter: https://itt-leipzig.de/ [23.02.2023].

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

schatztests eingesetzt.[135] Die Wortschatztests beruhen auf den Daten des bisher unveröffentlichten Herder/BYU-Korpus, mit dessen Hilfe Häufigkeitslisten für das Deutsche erstellt werden können. Da in der vorliegenden Studie nur schriftliche Daten der Teilnehmenden ausgewertet wurden, kann an dieser Stelle nicht weiter auf die in diesem Abschnitt aufgeführten Tests eingegangen werden. Für zukünftige Analysen wäre es höchst wünschenswert, die Daten der Lernenden für die verschiedenen Modalitäten und für rezeptive sowie produktive Kontexte systematisch zusammenzubringen.

Im zweiten Teil der Studie wurde die experimentelle Manipulation des Unterrichts fortgeführt. In einer der beiden Gruppen wurde ein Schwerpunkt auf Syntax und in der anderen Gruppe auf Semantik gelegt. Das Curriculum in beiden Studienphasen war jeweils als Spiralcurriculum angelegt, was vor allem bedeutete, dass bestimmte Inhalte (vor allem die Inhalte der Schwerpunktsetzungen) immer wieder auf komplexeren Ebenen im Unterricht behandelt wurden. Im Zeitraum der zweiten Studienphase wurde im Unterricht ein größerer Fokus auf rezeptive Fertigkeiten gelegt.[136] In dieser Phase wurden erneut drei anatomische MRT-Scans und zwei funktionelle fMRT-Scans auf Deutsch durchgeführt. Einige der behavioralen Tests wurden wiederholt, um Daten im Längsschnitt vergleichen zu können. Aus der zweiten Studienphase gehen lediglich Daten der beiden Teilnehmenden TN024 und TN025 im Rahmen der Einzelfallanalysen, die sich über die gesamte Studiendauer erstrecken, in die Analysen der vorliegenden Arbeit ein.

135 Dieses EIT ist bis jetzt nicht veröffentlicht. Es sind zwei Masterarbeiten im Zusammenhang mit dem EIT des ITT entstanden. Schleicher (2018) beschäftigt sich mit dem EIT als Instrument zur Erfassung grammatischer Kompetenz, während Röß (2019) mithilfe des EIT Zusammenhänge zwischen sozialer Integration und dem Erwerb des Deutsch als Zweitsprache untersucht. Weiterführende Informationen und Literaturhinweise zu den Wortschatztests finden sich auf den Seiten des ITT unter: https://itt-leipzig.de/wortschatztests#1505164286869–5364d3b3–8c9f [23.02.2023].

136 Diese Entscheidung wurde aus zwei Gründen getroffen. Einerseits legten die Prüfungsergebnisse der Teilnehmenden sowohl bei den Goethe-Zertifikatsprüfungen als auch bei den Prüfungen des ITT eine signifikante Differenz zwischen rezeptiven und produktiven Fähigkeiten nahe, wobei die produktiven die rezeptiven Fähigkeiten deutlich überstiegen. Andererseits deuten eine Reihe von Forschungsergebnissen der letzten Jahre darauf hin, dass diese Differenz ein häufig beobachtetes, sprachübergreifendes Phänomen ist, das nur durch eine signifikante Erhöhung des sprachlichen Inputs ab einer bestimmten Niveaustufe ausgeglichen werden kann (vgl. hierzu vor allem Tschirner 2016, der sich in seinen Ausführungen auf die Auswertung der Daten von über 3000 Studierenden an amerikanischen Universitäten stützt).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

3.1.4 Materialzugang und Messpunkte

Eine größere Anzahl an Messpunkten, die dann ein sehr dichtes Datenset konstituiert, kann die Ergebnisse präziser und reliabler machen. Außerdem kann individuelle Entwicklung genauer nachvollzogen werden (vgl. van Dijk / Verspoor / Lowie 2011: 59; vgl. Barkaoui 2014; vgl. Larsen-Freemann 2006). Barkaoui (2014) geht davon aus, dass es in longitudinalen Studiendesigns mindestens drei Messpunkte geben muss (vgl. Barkaoui 2014: 69). Bei der Auswahl der Messpunkte muss darüber hinaus ein Gleichgewicht gefunden werden zwischen aus der Theorie begründeten und daher sinnvollen Messintervallen und zeitlichen Messintervallen, die die Teilnehmenden einer Studie nicht überfordern und zu ‚Müdigkeit' bei diesen führen (vgl. ebd.).

Während der beiden Studienphasen wurde eine Vielzahl unterschiedlicher Daten erhoben. Im Rahmen der Erhebung von Sprachdaten der Teilnehmenden wurden insgesamt ca. 1250 schriftliche und etwa 120 mündliche Texte der Lernenden erfasst bzw. erhoben. Die schriftlichen Texte wurden transkribiert und im txt-Format sowie in xml-Dateien zur weiteren Analyse gespeichert. Die mündlichen Texte wurden nach den cGAT-Konventionen in EXMARaLDA transkribiert (vgl. Selting et al. 2009; vgl. Schmidt et al. 2015). Die statistischen Analysen wurden für diese Studie in Excel, IBM SPSS 26 sowie in JASP 0.16 durchgeführt. Mithilfe derselben Programme wurden ebenfalls die Abbildungen und Tabellen erstellt. In der vorliegenden Studie wurde nur ein Teil der vorhandenen Daten ausgewertet. Für die Analysen auf Gruppenebene wurden letztlich drei Zeitpunkte bzw. die von den Teilnehmenden produzierten (schriftlichen) Texte aus drei Sprachprüfungen gewählt, die während der ersten Studienphase im Abstand von etwa zwei Monaten durchgeführt wurden. Für die Einzelfallstudie mit den Lernenden TN024 und TN025, die sich über die gesamte Studiendauer erstreckt, wurden insgesamt 24 Zeitpunkte ausgewählt, die in der vorliegenden Studie ausgewertet wurden. Weiterhin ist darauf hinzuweisen, dass in der vorliegenden Untersuchung ausschließlich schriftliche Sprachdaten in die Analysen eingeflossen sind. Mündliche Vergleichsdaten der Teilnehmenden liegen vor, wurden aber im Rahmen dieser Arbeit nicht in die Analysen miteinbezogen. Hier in der Zukunft vergleichende Analysen durchzuführen, ist ein Desiderat für weitere Forschungsarbeiten. Bei der Analyse der schriftlichen Arbeiten wurden die jeweils zugrundeliegenden Aufgabenstellungen in die Analyse miteingeschlossen (vgl. dazu Kapitel 3.2.3 für die Gruppendaten und Kapitel 3.3.2 für die Einzelfallanalysen). Hier besteht durchaus die Möglichkeit, dass die Lernenden Textteile aus der Aufgabenstellung in ihre eigenen Textproduktionen einbauten bzw. diese direkt übernahmen. In der vorliegenden Studie wurden in Anlehnung an das Verfahren der Profilanalyse nach Grießhaber (vgl. Kapitel 2.4.3) keine Daten aufgrund systematischer Übernahmen aus der Analyse ausgeschlossen. Portmann (2020) verdeutlicht in diesem Zusammenhang jedoch die Relevanz einer genauen Analyse übernommener Strukturen und wie diese Ergebnisse unter Umständen verzerren können, etwa indem sprachlichen Produktionen bei einer Profilanalyse zu hohe Niveaustufen zugeordnet werden.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Ebenso schließt etwa Czinglar (2014) eine Reihe von Strukturen grundsätzlich aus den Analysen aus, da diese als formelhafte Chunks klassifiziert werden.[137] Czinglar bezeichnet mit dem Begriff Chunk „eine Ausdruckssequenz, die zusammen mit ihrer Bedeutung bzw. Funktion als Ganzes im Langzeitgedächtnis gespeichert und holistisch abgerufen bzw. verarbeitet wird" (ebd.: 112). In ihrer Analyse schließt Czinglar alle Kombinationen einer „finiten Verbform mit einem oder mehreren anderen Elementen" (ebd.: 114) als Chunks aus, wenn diese in ihrem Korpus mehr als 30-mal auftreten, um damit eine Verzerrung der beobachteten Verbstellungsmuster zu vermeiden (vgl. ebd.). Sie macht ebenfalls darauf aufmerksam, dass letztlich weder zu entscheiden ist, ob eine Struktur von Lernenden holistisch, d. h. als Chunk, abgespeichert ist oder nicht, noch, wann Chunks als solche aufgebrochen und in ihrer Struktur analysiert werden (vgl. ebd.). Aufgrund dieser grundsätzlichen Unsicherheit bei der Auswahl von Chunks wurde auf diesen weiteren Analyseschritt, nämlich produzierte Strukturen mit finiten Verbteilen aus der Analyse der Verbstellungsmuster auszuschließen, aus Kapazitätsgründen verzichtet.[138] Dies ist als weitere Einschränkung der Ergebnisse dieser Studie mitzudenken, da nicht auszuschließen ist, dass dies in Einzelfällen die erreichte Niveaustufe verzerrte.

Die Erfassung der Daten erfolgte im Rahmen des Unterrichts in einem möglichst natürlichen Kontext und ohne die Durchführung zusätzlicher Experimente (vgl. Schramm 2018: 35; Fandrych / Wallner 2022).[139] Dadurch entstammen die Aufgabenstellungen jeweils den konkreten Unterrichtskontexten. Dazu kommen die Textproduktionen, die in den Sprachtests im Laufe der Studie erhoben werden. Die erfassten und erhobenen Texte müssen folglich über einen gewissen Zeitraum in Beziehung zueinander gesetzt werden. Neben direkten Textübernahmen von Strukturen gerade bei beginnenden Schreibenden (vgl. Portmann 2020: 77) könnten darüber hinaus Aufgabeneffekte deutlich stärker ins Gewicht

137 Czinglar (2014) macht sehr transparent, welche Formen ausgeschlossen wurden und widmet den zugrundeliegenden Ausschlussentscheidungen ein ganzes Kapitel (vgl. Czinglar 2014: 112–120). Dies stellt ein überaus begrüßenswertes methodisches Vorgehen dar, lassen sich so doch Entscheidungen detailliert nachvollziehen. Prinzipiell ist in diesem Kontext die Frage interessant, inwiefern der Einschluss oder Ausschluss von Chunks einerseits gerechtfertigt ist und andererseits, wie solche Entscheidungen tatsächlich die beobachteten Verbstellungsmuster und damit unter Umständen die rekonstruierten Erwerbssequenzen beeinflussen können.

138 Bei Stichprobenanalysen wurden in der vorliegenden Arbeit nur wenige direkte wörtliche Übernahmen identifiziert, die gleichzeitig keinen wesentlichen Einfluss auf die vergebenen Niveaustufen zu haben scheinen.

139 Bei den Texten handelt es sich entweder um Schreibaufgaben, die im Unterricht von allen Lernenden unter vergleichbaren Umständen verfasst wurden oder um Prüfungsaufgaben, die während der durchgeführten Sprachprüfungen geschrieben wurden. Zur Textauswahl und zu den Aufgabenstellungen vgl. Kapitel 3.2.3 für die Gruppendaten und Kapitel 3.3.2 für die Einzelfallanalysen.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

fallen (vgl. zu möglichen Aufgabeneffekten z.B. Alexopoulou et al. 2017; vgl. Michel et al. 2019).[140]

Zusätzlich kann davon ausgegangen werden, dass Lernende im ‚schulischen Kontext' „nur einen geringen Teil ihres sprachlichen Repertoires nutzen" (Dirim / Krehut 2020: 535) und so nur teilweise spontaner Sprachgebrauch erfasst werden kann, der darüber hinaus nur einen Teil der tatsächlich zur Verfügung stehenden sprachlichen Ressourcen abbildet (vgl. ebd.: 536).[141] Die erfassten Daten illustrieren damit unter Umständen eher eine sprachlich korrigierte Form der tatsächlichen Sprachleistungen der Lernenden und nicht deren ungesteuerten Sprachgebrauch (vgl. ebd.: 546). Hinzu kommt, dass schriftliche Produktionen unter Umständen nur beschränkt Auskunft über die zugrundeliegende zweitsprachliche Kompetenz Auskunft geben können (vgl. Knappik / Dirim 2018: 169).

Veränderungen bzw. Entwicklungen müssen sich nicht unbedingt in den gemessenen Intervallen vollziehen. Es ist durchaus denkbar, dass Entwicklungen aufgrund der gewählten Messintervalle nicht beobachtet werden können. Trotzdem ist es naheliegend, dass es Momente und Episoden gibt, in denen Entwicklungen beobachtet werden können (etwa im Zusammenhang mit der Instruktion einer Struktur im Unterricht) (vgl. Baba / Nitta 2014: 3).

3.1.5 Experimentelle Manipulation

Der gesamten Studie lag eine experimentelle Manipulation zugrunde, die Aufschluss darüber geben sollte, ob eine unterschiedliche Instruktion hinsichtlich semantischer bzw. syntaktischer Inhalte verschiedene neuronale Repräsentationen im Gehirn zur Folge hat und ob sich Unterschiede zwischen den Teilnehmenden feststellen lassen, je nachdem, welche Instruktion sie erhalten haben. Diese Forschungsperspektive der Gesamtstudie ergab sich aufgrund der in der Einleitung (Kapitel 1.3) präsentierten neurowissenschaftlichen Desiderata. Für die Durchführung der Studie bedeutet dies, dass die Lernenden jeweils zufällig einer Experimentalgruppe zugeordnet wurden. Die Studiengruppe war in sechs Kurse aufgeteilt, von denen in drei Kursen explizit ein Schwerpunkt auf syntaktische Kursinhalte und in drei Kursen ein Schwerpunkt auf semantische Kursinhalte gelegt wurde und syntaktische Inhalte eher implizit vermittelt wurden.

140 In Studien zur Schreibentwicklung werden oft dieselben Aufgaben mehrfach eingesetzt, was Entwicklungen klarer nachvollziehbar macht. Im SpraStu-Projekt wird für die Erstellung des DISKO-Korpus etwa mehrfach dieselbe TestDaF-Prüfung zum Schreiben durchgeführt (vgl. Muntschick et al. 2020: 8; Wisniewski / Möhring 2021).

141 Dirim / Krehut (2020) nehmen Kinder und Jugendliche in den Blick. Die getroffenen Aussagen scheinen ebenso auf erwachsene Lernende in institutionellen Kontexten bzw. in gesteuerten Erwerbskontexten übertragbar.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Die zu vermittelnden Inhalte und die in diesem Zusammenhang ausgewählten Übungen wurden den Lehrkräften jeweils in den Rahmenplänen für die jeweilige Unterrichtswoche zur Verfügung gestellt (vgl. dazu Kapitel 3.1.3). Die beiden Schwerpunkte hatten insgesamt etwa einen Umfang von 60 UE, was für die verschiedenen Kurse einem Anteil von 10–12,5 % des Gesamtunterrichtsvolumens entspricht.[142] Für die hier vorliegende Studie ist die experimentelle Manipulation insofern relevant, als dass mögliche Auswirkungen der Gruppeneinteilung bei der Interpretation der Ergebnisse berücksichtigt werden müssen (vgl. zu diesem Punkt Kapitel 3.2). Schematisch stellt sich diese Manipulation in den beiden Studienphasen wie folgt dar:

Phase 1 (n=65)[143]:

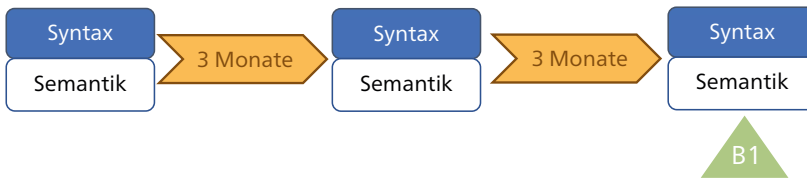


Abbildung 1: Studienphase 1 von Mai 2016–Februar 2017 (schematisch)

Phase 2 (n=42)

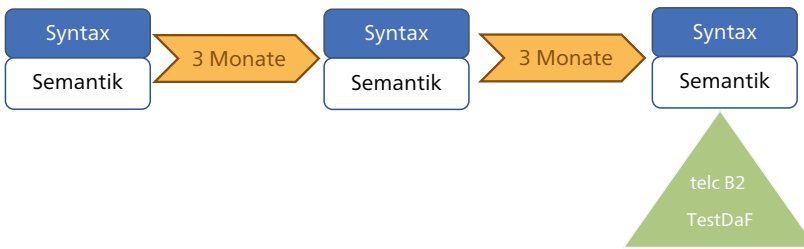


Abbildung 2: Studienphase 2 von März 2017–September 2017 (schematisch)

142 Diese Zahl ergibt sich aus dem Verhältnis von Unterrichtsstunden, die explizit als Syntax- bzw. Semantikstunden gekennzeichnet waren, zu den möglichen Gesamtstunden, die Lernende im Rahmen des Unterrichts hätten besuchen können. Dabei werden die Unterrichtsstunden über alle drei (bzw. sechs) Kurse gemittelt.

143 Die Intervalle von drei Monaten ergeben sich aus den zu Beginn der Studie angesetzten Scanintervallen innerhalb der Studie. Diese konnten nicht immer exakt eingehalten werden, daher hat die Abbildung vor allem schematischen Charakter. Die Zahl n=65 ergibt sich an dieser Stelle aus den Teilnehmenden, die die erste Studienphase abgeschlossen haben. Teilnehmende, die den Kurs nach kurzer Zeit abgebrochen haben, werden hier nicht mitaufgeführt (vgl. dazu Kapitel 3.1.6.3).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

In Phase 1 bestand die Schwerpunktsetzung im Bereich Syntax darin, dass einerseits bestimmte Inhalte extensiver mit den Teilnehmenden thematisiert und andererseits deutlich früher als in der Progression der Lehrwerke vorgesehen im Unterricht eingeführt wurden. Inhalte, die verstärkt unterrichtet wurden, sind die für das Deutsche spezifische Satzklammer oder linkserweiternde Strukturen (vgl. Fandrych / Thurmair 2016). Zu Beginn des A2-Kursteils wurden subordinierende Nebensätze und Relativsätze eingeführt, im Gegensatz zu den meisten Lehrwerken, die diese Inhalte erst am Ende des A2-Niveaus oder zu Beginn des B1-Niveaus einführen. Die Vermittlung im Unterricht fand explizit statt, wobei darauf geachtet wurde, dass die Grammatik- bzw. die Syntaxvermittlung in kommunikative Kontexte eingebettet war und von den Lernenden nicht als Selbstzweck aufgefasst wurde. Dies geschah in dem Wissen, dass deklaratives syntaktisches Wissen „nicht umstandslos in praktischer Sprachkompetenz umsetzbar ist" (Fandrych 2013: 171).[144] Es wurde folglich versucht, den Lernenden deklaratives Grammatikwissen so zur Verfügung zu stellen, dass es gleichzeitig als ein Werkzeug zur Realisierung größerer Autonomie dienen konnte. Dies galt sowohl für Situationen im Alltagsleben der Teilnehmenden als auch für die Lehr- und Lernsituationen, in denen die Lernenden sich befanden (vgl. hierzu Fandrych 2000; Fandrych 2013: 170).

Im semantischen Schwerpunkt wurden in der ersten Studienphase vor allem Bedeutung und Verwendung von Modalverben und Strategien zur Wortschatzerweiterung und zum Erschließen unbekannter Wörter aus dem Kontext behandelt. In den Gruppen mit semantischen Schwerpunkten gab es ebenfalls immer wieder Phasen, die auf grammatische Strukturen fokussierten, um zu verhindern, dass bestimmte Strukturen des Deutschen von den Lernenden komplett übersehen werden (vgl. Fandrych 2013: 170). Hier wurde jedoch Wert auf eine eher implizite Vermittlung der Strukturen gelegt.

In der zweiten Studienphase wurde der Anteil der experimentellen Manipulation am Gesamtumfang des Unterrichts deutlich auf etwa 18 % des Gesamtunterrichts erhöht. Einerseits wurde in der Syntaxgruppe verstärkt explizit instruiert, im Gegensatz zur Semantikgruppe, in der die syntaktischen Inhalte eher implizit und mit einer Konzentration auf Input eingeführt wurden. Inhaltliche Schwerpunkte wurden in der Syntaxgruppe auf den Aufbau grammatischer Lese- und Hörmuster (vgl. Lauterbach 2005), auf die Realisierung von Sätzen mit Objekten in Erststellung und auf die Vermittlung syntaktischer Strukturen von verschie-

144 Aus neurowissenschaftlicher Sicht wurde bei der Konzeption der experimentellen Manipulation an das Modell des deklarativen / prozeduralen Gedächtnis von Ullman angeschlossen, das am ehesten mit dem Modell des *weak interface* bei N. Ellis kompatibel ist (vgl. Ellis 2005 und 2007; Ullmann / Lovelett 2018). In dieser Modellbildung wird davon ausgegangen, dass explizites und implizites Wissen beim Sprachenlernen und beim Sprachgebrauch auf verschiedenen Systemen bzw. Netzwerken im Gehirn beruht (vgl. Ullman / Lovelett 2018: 45).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

denen Textsorten gelegt. Diese Schwerpunktsetzung durfte dabei nicht außer Acht lassen, dass eine erhöhte Aufmerksamkeit gegenüber syntaktischen Strukturen zu einem expliziteren Wissen um diese führen kann (vgl. DeKeyser 2015; Rosa / Leow 2004).[145] In der Semantikgruppe wurden syntaktische Inhalte eher implizit vermittelt. Gleichzeitig wurden inhaltliche Schwerpunkte auf Modalpartikeln und die semantische Differenz verschiedener Präfixe und Suffixe gelegt. Textsorten wurden vor allem in Bezug auf ihre „kommunikationssituativen und funktionalen Charakteristika" (Fandrych / Thurmair 2011) behandelt.

Die möglichen Effekte dieser experimentellen Manipulation bzw. unterrichtlichen Intervention wurden bei der Auswertung der Daten jeweils auf Gruppenebene statistisch geprüft. Dies geschah zu den drei Zeitpunkten, an denen der A1/A2-Test, der A2-Test und das Goethe-Zertifikat B1 durchgeführt wurden und anhand der aus den Texten jeweils berechneten Oberflächenmaßen (vgl. Kapitel 3.2.5 für die Operationalisierung der verwendeten Maße). Für die zweite Studienphase wurde in der vorliegenden Arbeit kein Vergleich zwischen den beiden Experimentalgruppen durchgeführt, da die drei ausgewählten Zeitpunkte innerhalb der ersten sechs Monate der Studie liegen. Dies bedeutet, dass keine abschließenden Aussagen zur Wirksamkeit der experimentellen Manipulation getroffen werden können, die von den in der zweiten Studienphase erhobenen Daten ausgehen. Die beiden Lernenden in den Einzelfallstudien sind Geschwister und kommen beide aus der semantischen Experimentalgruppe, um im Kontext dieser Studie eine größtmögliche Vergleichbarkeit herzustellen.

3.1.6 Limitationen der Studie

In diesem Abschnitt wird versucht, Forschungsprozesse und methodische Einschränkungen, die die Gültigkeit der vorgelegten Ergebnisse beeinflussen können, möglichst transparent zu machen. Zusätzlich werden die Ergebnisse der vorliegenden Studie in Kapitel 5 kritisch diskutiert und mögliche Einschränkungen besprochen. Ein Anspruch auf Vollständigkeit der möglichen Limitationen kann an beiden Stellen nicht erhoben werden. Zunächst ist darauf hinzuweisen, dass sich Einschränkungen daraus ergeben, dass trotz einer sehr extensiven Datensammlung (vgl. dazu Kapitel 3.1.4) bestimmte Daten in der Studie nicht erhoben wurden. Dies bezieht sich, wie oben beschrieben, etwa auf persönliche Daten im Zusammenhang mit der Fluchtgeschichte der Teilnehmenden. Ergebnisse in den Kontext persönlicher Hintergrunderfahrungen der Lernenden zu setzen ist folglich nicht möglich (vgl. Kapitel 3.1.2). Im Folgenden werden die Position des Forschenden deutlich gemacht (vgl. Kapitel 3.1.6.1), die Zugänglichkeit der Daten kritisch in den Blick genommen (vgl. Kapitel 3.1.6.2)

145 Dies wiederum könnte Prozesse bei der Verarbeitung im deklarativen Gedächtnis befördern.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

und die Abbrecherinnen und Abbrecher als Faktor für mögliche Verzerrungen analysiert (vgl. Kapitel 3.1.6.3).

3.1.6.1 Die Position des Forschenden

Während der Durchführung der Studie war ich als Studienkoordinator am MPI CBS tätig und mit allen Teilnehmenden bekannt. Diese Zusammenarbeit war zum Teil sehr persönlich, was dazu führte, dass ich Hintergrundwissen über die Lebenssituation und die persönlichen Umstände der Teilnehmenden erwarb, welches nicht im Rahmen der Studie erhoben wurde. In diesen Fällen war es mir wichtig, dass ich dieses Wissen aus Datenschutzgründen nicht in die vorliegende Dissertation einfließen ließ und ich hoffe, dass dies gelungen ist. Es kann nicht völlig ausgeschlossen werden, dass Wissen in die Analysen und in die Interpretationen der Ergebnisse eingeflossen ist, welches nicht im Rahmen der offiziellen Erhebungen gesammelt wurde. Ich war darüber hinaus für die Konzeption des Curriculums der einzelnen Kurse und die Realisierung der experimentellen Manipulation sowie die Zusammenarbeit mit den Lehrkräften verantwortlich. Diese Zusammenarbeit umfasste Besprechungen mit den Lehrkräften und die Koordination der Arbeit der verschiedenen studentischen Hilfskräfte, die im Projekt angestellt waren. Einen weiteren Arbeitsbereich bildete die Koordination zwischen Max-Planck-Institut und Herder-Institut und die Zusammenarbeit mit den verschiedenen Testanbietenden (Goethe-Institut und ITT). Es ist aufgrund dieser exponierten Position des Forschenden zu hinterfragen, welchen Einfluss diese Position auf den Verlauf des Sprachkurses und damit auf die vorliegenden Ergebnisse haben könnte. Hier ist sicherlich anzumerken, dass die gesamte Studie unter großem Zeitdruck für alle Beteiligten durchgeführt wurde, da der Spracherwerb ‚live' ablief und die Teilnehmenden wiederum großem Druck ausgesetzt waren, so schnell wie möglich bestimmte sprachliche Fertigkeiten und Niveaustufen zu erreichen, um dann Zertifikate vorweisen zu können oder weitergehende Ausbildungsschritte angehen zu können. Gleichzeitig war jedoch das Tempo des Kurses zum Teil zu hoch angesetzt und musste in einigen Gruppen immer wieder angepasst werden. Das war eine Herausforderung, da gleichzeitig sichergestellt werden musste, dass die einzelnen Unterrichtsgruppen möglichst vergleichbar blieben. Bei der Konzeption der vorliegenden Arbeit wurden keine qualitativen Erhebungsmethoden eingeplant, bei denen die persönliche Beziehung zwischen Teilnehmenden und Forschenden eine kritische Rolle spielen würde (Interviews, teilnehmende Beobachtung etc.). Im Gegenteil, es wurde versucht, aus den Sprachproduktionen der Teilnehmenden Werte zu extrahieren, die einen möglichst ‚objektiven' Einblick in die Zweitsprachentwicklung dieser Teilnehmenden vermitteln können. Zur Bewertung bzw. zur Operationalisierung dieser Entwicklung wurden darüber hinaus nur direkt beobachtbare Werte und Testergebnisse herangezogen und keine Bewertungen der Lehrkräfte.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

3.1.6.2 Zugänglichkeit der Daten

Um die der Arbeit zugrundeliegenden Daten nachvollziehbar und die Ergebnisse überprüfbar zu machen, finden sich die Ergebnisse aller Profilanalysen und alle herausgearbeiteten Oberflächenmaßen im Anhang. Ebenso finden sich dort umfangreiche Tabellen, in denen Hintergrunddaten zu einzelnen Lernenden verfügbar sind. Die Lernertexte selbst können bis jetzt auf Anfrage zur Verfügung gestellt werden. Eine Veröffentlichung als eigenständiges und aufbereitetes Korpus ist geplant und ein im Sinne der FAIR-Prinzipien aufbereitetes, durchsuchbares und öffentlich zugängliches Korpus im Rahmen des größeren Forschungsprojektes am MPI CBS und am HI würde die Nachvollziehbarkeit und die Transparenz der durchgeführten Analysen entscheidend erhöhen.[146] Zudem existieren bereits zahlreiche, umfangreich aufbereitete und zugängliche Lernerkorpora wie die Falko-Lernerkorpusfamilie (vgl. u. a. Lüdeling et al. 2008) mit Korpora wie dem KanDel-Korpus (vgl. Vyatkina 2016), das trilinguale MERLIN-Lernerkorpus (vgl. Abel et al. 2014; vgl. MERLIN-Projekt 2014) oder DISKO-Lernerkorpus (vgl. Muntschick et al. 2020; Wisniewski et al. 2022a).

3.1.6.3 Abbrecherinnen und Abbrecher

Eine grundlegende Herausforderung in longitudinalen Studien ist der Schwund bzw. die Fluktuation (*attrition*) von Teilnehmenden während des Studienverlaufs (vgl. Little / Rubin 2002: 7). Es ist nicht ungewöhnlich, dass nicht alle Teilnehmenden zu jedem Zeitpunkt an jedem Test teilnehmen oder in einem Test nicht alle Aufgaben beantworten (vgl. Pichette et al. 2015). Daher kommt es immer wieder zu fehlenden Daten (*missing values*). Zusätzlich brechen manche Teilnehmende longitudinale Studien komplett ab und diese Gruppe der Abbrecherinnen und Abbrecher (*Dropouts*) kann wiederum in den statistischen Berechnungen eine große Rolle spielen, da nicht evident ist, ob es Eigenschaften gibt, die diese Gruppe im Unterschied zum Rest des Samples miteinander verbindet (vgl. Barkaoui 2014: 73). Dabei kann es sowohl dazu kommen, dass die Ergebnisse durch die Abbrecherinnen und Abbrecher deutlich verzerrt sein könnten, als auch dazu, dass dies nicht der Fall ist.[147] Während die Zahl der

146 FAIR steht für ‚Findable, Accessible, Interoperable, and Re-usable'. Forschungsdaten sollen leicht auffindbar und zugänglich sein und zusätzlich mit unterschiedlichen Werkzeugen bearbeitbar und in gängigen Datenformaten vorliegen. Der letzte Punkt bezieht sich auf die gewünschte Wiederverwendbarkeit der vorliegenden Daten, die im Kontext möglicher Replikationsstudien von großer Bedeutung sind. Die FAIR-Prinzipien bilden einen wichtigen Aspekt beim Aufbau der nationalen Forschungsdateninfrastruktur (NFDI; vgl. Brünger-Weilandt et al. 2019), über die ein zentrales und zugängliches Forschungsdatenmanagement aufgebaut und zur Verfügung gestellt werden soll.

147 Bell et al. (2013) weisen im Kontext klinischer Studien auf den Umstand hin, dass ein differentieller *Dropout*, d.h. ein *Dropout*, der sich in ungleichmäßiger

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Dropouts in vielen Studien zumindest berichtet wird, wird nur in wenigen Studien genauer analysiert, durch welche Eigenschaften diese *Dropouts* sich im Vergleich zum Rest der Stichprobe auszeichnen (vgl. Rees / Klapper 2008).[148] Dies kann unter anderem damit zusammenhängen, dass für Teilnehmende, die sehr früh abbrechen, nur wenige Daten verfügbar sind, die für Vergleiche genutzt werden können. In dieser Arbeit wurde grundsätzlich darauf verzichtet, fehlende Daten mithilfe komplexer Verfahren zu schätzen bzw. fehlende Werte zu ersetzen (für eine umfassende Einführung in verschiedene Techniken für den Umgang mit fehlenden Daten vgl. Little / Rubin 2002; vgl. Baraldi / Enders 2010). In vielen Fällen kam es dazu, dass etwa ein Testergebnis oder eine Textproduktion fehlte, die folgende aber wieder vorhanden war – solche Datensätze wurden bei der Analyse auf Gruppenebene nicht berücksichtigt. Vielmehr wurden zur Beantwortung der Forschungsfragen jeweils nur vollständige Datensätze herangezogen, das heißt, dass Datensätze, bei denen nicht alle untersuchten Variablen vorhanden waren, aus den Analysen ausgeschlossen wurden (vgl. Pichette et al. 2015: 155). Aus diesem Grund werden in der vorliegenden Arbeit für jede Studienphase und für jede Forschungsfrage jeweils die ausgewählte Stichprobe der Teilnehmenden mit der gesamten Stichprobe aller Teilnehmenden (inklusive der jeweils ausgewählten Teilstichproben) verglichen, um mögliche Verzerrungen besser einschätzen zu können. Darüber hinaus werden in den Fällen, in denen Teilnehmende aufgrund unvollständiger Datensätze ausgeschlossen werden, diese Entscheidungen zum einen begründet und zum anderen kontextualisiert werden. Dabei geht es nicht darum, Gründe für die Unterschiede zwischen den verschiedenen Gruppen herauszuarbeiten oder mögliche Gründe für den Abbruch der Studie zu rekonstruieren. Es wird vielmehr davon ausgegangen, dass diese Gründe immer multikausal sind und unter Umständen durch die in dieser Studie erhobenen Variablen gar nicht abgedeckt werden können. Ziel dieses Kapitels ist eine vergleichende Darstellung der sich durch die Abbrüche ergebenden Gruppen. Es werden daher folgende Vergleiche eine Rolle spielen:

1. *Dropouts* nach Abschluss der Rekrutierung im Vergleich zur gesamten Stichprobe (N=94) und während des ersten Teils der Studie. Hier sind nur wenige Daten vorhanden und daher kaum Vergleiche möglich.

Weise über verschiedene Experimentalgruppen verteilt, zu einer Verzerrung der Ergebnisse führen könnte, dies aber nicht notwendigerweise tut und gleichzeitig ein gleichmäßig über die Gruppen verteilter *Dropout* ebenso zu Verzerrungen führen könnte.

148 Pichette et al. (2015) zeichnen hier ein deutlich negativeres Bild. In einer Review-Studie zur Exklusion von Teilnehmenden mit fehlenden Daten ermitteln die Forscher/-innen, dass nur in 3,5 % der von ihnen untersuchten Studien der Ausschluss von Teilnehmenden mit fehlenden Daten berichtet wird (Pichette et al. 2015: 155). Sie weisen zusätzlich darauf hin, dass es kein allgemein akzeptiertes, standardisiertes Verfahren zum Umgang mit fehlenden Daten gibt (vgl. ebd.: 156).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

2. Ein Vergleich zwischen allen Teilnehmenden, die B1 abgeschlossen haben (n=62) und den Teilnehmenden, die für die in dieser Studie untersuchte Stichprobe (n=44) ausgewählt wurden (vgl. Kapitel 3.2.1 zur Teilstichprobe für Forschungsfragen 1 und 2 und Kapitel 3.2.2 zu Unterschieden zwischen Teilstichprobe und ausgeschlossenem Datenset)
3. Ein Vergleich zwischen den Teilnehmenden, die für den zweiten Teil der Studie ausgewählt wurden (n=42) und den beiden für die Einzelfallanalysen ausgewählten Lernenden (n=2) (vgl. Kapitel 3.3.1 zur Beschreibung der beiden Lernenden TN024 und TN025)

Innerhalb der ersten Studienphase brachen insgesamt 29 Teilnehmende (~30,1 %) die Partizipation an der Studie ab.[149] Für 62,1 % (n=18) davon war dies innerhalb der ersten acht Wochen nach Unterrichtsbeginn der Fall, d.h. noch vor der A1-Prüfung oder dem zweiten MRT-Scan nach drei Monaten (Mittelwert = 10,2759; *Mdn* = 8; SD = 7,51140; Zeitangaben in Wochen). Das bedeutet gleichzeitig, dass für diese 18 Personen so gut wie keine Daten vorliegen. Bezogen auf das Alter lässt sich feststellen, dass die Abbrecherinnen und Abbrecher im Schnitt 24,76 Jahre alt waren (*Mdn* = 25; SD = 4,904).[150] Hinsichtlich der Gruppeneinteilung der experimentellen Manipulation brachen 16 Personen aus Syntaxkursen und 13 Personen aus Semantikursen die Studie ab. 23 (79,3 %) der Dropouts waren männlich und 6 (20,7 %) weiblich. Der Mittelwert des Alters der Teilnehmenden, die die Studie weiter fortsetzten, ist 24,26 (*Mdn* = 24; SD = 4,725). Von diesen Teilnehmenden besuchten 32 Personen die Kurse mit syntaktischem und 33 Personen die Kurse mit semantischem Schwerpunkt (Mittelwert = 0.51; SD = 0,504). 51 (78,6 %) der Teilnehmenden, die den Kurs weiter fortsetzten, gaben ihr Geschlecht als männlich und 14 (21,5 %) als weiblich an. Alle Kenndaten sind in Tabelle 5 zusammengefasst:

	Dropouts (n=29)			Teilnehmende, die den Kurs fortsetzen (n=65)		
	Mittelwert		Standard-abweichung	Mittelwert	Median	Standard-abweichung
Alter	24,76	25	4,904	24,26	24	4,725
Experimentelle Manipulation	0,45	0	0,506	0,51	1	0,504
Geschlecht	0,21	0	0,412	0,22	0	0,414

Tabelle 5: Vergleich zwischen Dropouts und Teilnehmenden, die den Kurs in den ersten sechs Monaten besuchen.

149 Allgemeine Zahlen bezüglich der Abgänge aus Integrationskursen legen Cindark / Henrich (2019) vor. In den von ihnen untersuchten Integrationskursen berichten sie von einer *Dropout*-Quote von 34 % (vgl. Cindark / Henrich 2019: 21).

150 Das hier und im folgenden angegebene Alter bezieht sich jeweils auf das Alter zu Beginn der Studie. Die deskriptiven Daten verdeutlichen, dass sich die Gruppe der *Dropouts* in der Anfangsphase nicht in relevanter Weise von den Teilnehmenden unterscheidet, die die Studie fortsetzten. Einschränkend ist zu beachten, dass für diese erste Gruppe an Abbrechenden und Abbrechern nur sehr wenige belastbare Daten zur Verfügung stehen. Die deskriptiven Daten dienen der Überprüfung der Kerndaten, dem die Studie gestartet wurde und den Geburtsdaten der Teilnehmenden (vgl. Kapitel 3.1.2 und 3.1.3).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Geschlecht ergab ebenfalls keine signifikanten Unterschiede zwischen diesen beiden Gruppen:

- Zur Überprüfung der Altersunterschiede zwischen den beiden Gruppen wurde ein t-test mit zusätzlichem Bootstrapping durchgeführt. Der Altersdurchschnitt der *Dropouts* (n=29) (M = 24,76, SE = 0,57) liegt leicht über dem Altersdurchschnitt der Teilnehmenden, die den Kurs weiterführen (n=65) (M = 24,26, SE = 0,91). Dieser Unterschied von –0,5, BCa 95 %KI [–2,617, 1,623] ist nicht signifikant t(92) = –0,466, p = 0,643.[151] Die Effektgröße d beträgt 0,11.[152] Die Ergebnisse (d.h. die Konfidenzintervalle) des Bootstraps beruhen auf jeweils 1000 Bootstrap-Stichproben.
- Zur Überprüfung möglicher Gruppenunterschiede in Bezug auf die Gruppeneinteilung der experimentellen Manipulation wurde ebenfalls ein t-test mit zusätzlichem Bootstrapping durchgeführt. Hier brachen etwas mehr Personen aus den Syntaxkursen als aus den Semantikkursen die Studie ab. Der hier gefundene Unterschied von 0,059 BCa 95 % KI [–0,168; 0,284] ist ebenfalls nicht signifikant t(92) = 0,526, p = 0,601. Die Effektgröße d beträgt –0,119.
- Die Differenz bei der Variable ‚Geschlecht' ist zwischen den Dropouts und den übrigen Teilnehmenden der Studie bei einem Unterschied von 0,008 BCa 95 % [–0,190; –0,199] und t(92) = 0,092, p = 0,927 nicht signifikant. Hierbei ergibt sich eine Effektgröße d von –0,024.

151 BCa 95 % KI steht hier für *Bias corrected accelerated 95 % Confidence Interval*. Biaskorrigierte mit Bootstrap ermittelte Konfidenzintervalle (BCa 95 % CI) werden in eckigen Klammern angegeben (vgl. Field 2018: 364). Das Konfidenzintervall ist ein Bereich von Werten um eine berechnete Statistik herum. Es wird angenommen, dass der ‚wahre' Wert der Statistik in einem bestimmten Anteil der Stichproben (zum Beispiel bei 95 % der Stichproben, die aus den zugrundeliegenden Daten gezogen werden können) innerhalb dieses Bereichs liegt (vgl. Field 2018: 1011). Ein großes Konfidenzintervall deutet daher auf eine größere Unsicherheit bei der Schätzung der berechneten Statistik hin (vgl. Backhaus et al. 2018: 89). Ein 95 %-Konfidenzintervall bedeutet gleichzeitig, dass für einen bestimmten Anteil der Stichproben (5 %) das Konfidenzintervall den wahren Wert der Statistik nicht enthält (vgl. Field 2018: 1011).

152 Hier wurde aus den Mittelwerten beider Gruppen und der Standardabweichung der Gruppe, die den Kurs weitergeführt haben, Cohen's d berechnet (vgl. zur Berechnung von Effektstärken bei t-tests mit unabhängigen Stichproben Field (2018: 461)).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Alle drei Effektgrößen sind damit sehr gering und zusammen mit den nicht signifikanten Ergebnissen stützen sie hinsichtlich der hier ausgewählten Parameter die Schlussfolgerung, dass sich die Gruppe der Dropouts nicht signifikant von der Gruppe der Teilnehmenden unterscheidet, die die Studie fortsetzten.

3.1.7 Zusammenfassung der Grundlagen

In den vorangegangenen Abschnitten wurden die Teilnehmenden der Studie vorgestellt und der longitudinale Aufbau der Studie näher beschrieben. Die Studie wurde in zwei Phasen durchgeführt und erstreckte sich insgesamt von Mai 2016 bis September 2017. Die Teilnehmenden stammten als erwachsene Deutschlernende mit Arabisch als L1 aus einer relevanten Zielgruppe für den Kontext Deutsch als Zweitsprache, die bisher nur wenig im Fokus empirischer Forschung steht. Bei der Beschreibung der Teilnehmenden wurde deutlich, dass diese als Geflüchtete aus einer vulnerablen Zielgruppe stammen. Es ist durchaus möglich, dass traumatische Erfahrungen, die etwa durch den Krieg in den Heimatregionen der Teilnehmenden oder durch die Flucht nach Deutschland gemacht wurden, den Lernprozess beeinflussten. Daten zu möglichen psychologischen Erkrankungen wurden nicht erhoben, um die Teilnehmenden psychisch nicht zusätzlich zu belasten, und konnten daher nicht in die folgenden Analysen miteinbezogen werden. Die Zusammensetzung der Kurse innerhalb der vorliegenden Studie ist in einigen Aspekten mit der Zusammensetzung in Integrationskursen (vgl. Cindark / Hünlich 2019) vergleichbar (z. B. hinsichtlich der Geschlechterverteilung innerhalb der Kurse). Die Teilnehmenden dieser Studie sind etwas jünger als in den untersuchten Integrationskursen. Außerdem wurde die experimentelle Manipulation der Studie vorgestellt, in deren Rahmen die Teilnehmenden randomisiert zu Syntax- oder Semantikgruppen zugeteilt wurden, in denen bestimmte Inhalte jeweils unterschiedlich im Unterricht eingeführt und vermittelt wurden.

Im Kern der Studie stand ein Intensivsprachkurs, der an fünf Tagen pro Woche jeweils fünf Unterrichtseinheiten umfasste. Die Teilnehmenden hatten zu Beginn keine oder nur geringe Vorkenntnisse im Deutschen und der Kurs führte sie in der ersten Studienphase über das Niveau A1 bis zum Niveau B1 und in der zweiten Studienphase über das Niveau B2 bis zu den Anfängen von Niveau C1. Im letzten Abschnitt wurden einige Limitationen dieser Studie diskutiert, die sich vor allem auf die Aufbereitung der Daten, die Position des Forschenden und die Abbrecherinnen und Abbrecher und auf ihre möglichen Effekte hinsichtlich der erhobenen Daten innerhalb der Studie bezogen.

Aus der beschriebenen Gesamtstichprobe wurden zur Beantwortung der Forschungsfragen zwei Teilstichproben bzw. Subsamples gebildet, die in den nächsten beiden Kapiteln näher beschrieben werden (vgl. dazu Kapitel 3.2 zu den Gruppendaten und Kapitel 3.3 zu den Einzelfallanalysen). Auf Gruppen-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

ebene wurde eine Teilstichprobe aus 44 Teilnehmenden gebildet und für die vergleichende Einzelfallanalyse wurden zwei Geschwister ausgewählt, die denselben Kurs innerhalb der Studie besuchten.

3.2 Methodisches Vorgehen: Gruppendaten

In diesem Kapitel wird das methodische Vorgehen in Bezug auf die Auswertung der Gruppendaten vorgestellt. Dabei werden die einzelnen Methoden und Analysen nacheinander dargelegt. Die erhobenen Daten beziehen sich auf die Daten von 44 Teilnehmenden zu drei Zeitpunkten (vgl. zur hier verwendeten Teilstichprobe Kapitel 3.2.1 und 3.2.2) und auf Forschungsfragen 1, 2 und 3:

Forschungsfragen zur Entwicklung syntaktischer Kompetenz über die Zeit

Forschungsfrage 1: *Welche syntaktische(n) Erwerbsreihenfolge(n) ergibt (ergeben) sich für erwachsene Deutschlernende mit Arabisch als L1?*

Forschungsfrage 2: *Wie lässt sich die syntaktische Entwicklung der Teilnehmenden in den ersten sechs Monaten des Zweitsprachenlernens charakterisieren?*

Forschungsfrage zur experimentellen Manipulation und zur Instruktion

Forschungsfrage 3: *Lassen sich aufgrund der Instruktion bzw. der experimentellen Manipulation innerhalb der Studie (explizite Sprachvermittlung mit Schwerpunkt auf Syntax vs. implizite Vermittlung mit Schwerpunkt auf Semantik) Unterschiede zwischen den Teilnehmerinnen und Teilnehmern feststellen?*

Die in dieser Studie verwendete Kombination an Methoden soll helfen, zu verstehen, wann welche syntaktischen Strukturen geeignet sind, um zweitsprachliche Entwicklung zu beschreiben und auf welchen Sprachniveaus unter Umständen welche Konstrukte besonders aussagekräftig sind. Darüber hinaus wurden die Methoden so gewählt, dass sie intra- und interindividuelle Variabilität der Lernenden sichtbar machen können. Zur Analyse der Gruppendaten wurden zunächst für alle Texte Profilanalysen nach Grießhaber durchgeführt. Die Profilanalysen lieferten die jeweils erreichten Profilstufen und die in den einzelnen Texten produzierten Wortstellungsmuster. Außerdem wurden auf Grundlage der produzierten Texte syntaktische Oberflächenmaße herausgearbeitet, die den Analysen in diesem Teil zugrunde lagen (vgl. zur Auswahl der Texte Kapitel 3.2.3). Ausgehend von den in den Profilanalysen herausgearbeiteten Satzstrukturen innerhalb der produzierten Texte wurden Emergenzanalysen durchgeführt, bei denen verschiedene Emergenzkriterien geprüft wurden (vgl. Kapitel 3.2.4). Zuletzt wurden anhand der syntaktischen Oberflächenmaße longitudinale Clusteranalysen durchgeführt, um einerseits Gruppen in den Daten sichtbar zu machen und andererseits typische Entwicklungsverläufe zu identifizieren (vgl. Kapitel 3.2.5).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

3.2.1 Teilstichprobe für Forschungsfragen 1 und 2

Zur Beantwortung der ersten beiden Forschungsfragen wurde aus allen Teilnehmenden eine Teilstichprobe gebildet. Auswahlkriterium war hier, dass Daten zu drei ausgewählten Zeitpunkten aus der ersten Studienphase vorliegen (A1/2-Test, A2-Test und B1-Prüfung), um für diese Teilstichprobe Entwicklungslinien nachzeichnen zu können und um sicherzustellen, dass zu jedem Zeitpunkt dieselben Teilnehmenden in den Clusteranalysen analysiert wurden. Von den 65 Teilnehmenden, die bis zum Abschluss der ersten Studienphase (B1-Prüfung) an der Studie teilgenommen haben (29 Teilnehmende haben den Kurs vorzeitig abgebrochen; vgl. Kapitel 3.1.6.3 zu Abbrecherinnen und Abbrechern), stehen für 44 Teilnehmende vollständige Datensets zur Verfügung. Das heißt, dass die Daten von 21 Teilnehmenden, die die erste Studienphase abgeschlossen haben, in den folgenden Analysen nicht berücksichtigt wurden. Wie schon in Kapitel 3.1.6.3 zu den Abbrecherinnen und Abbrechern sollen hier die beiden Gruppen (Teilstichprobe und unvollständige und daher ausgeschlossene Datensets) miteinander verglichen werden, um zu evaluieren, inwiefern Unterschiede bestehen und welchen Einfluss diese Unterschiede auf die Ergebnisse der folgenden Analysen haben könnten. Es werden zunächst wie zuvor die Variablen ‚Alter', ‚Gruppenzugehörigkeit in der experimentellen Manipulation' und ‚Geschlecht' verglichen. Dann werden zusätzlich Variablen verglichen, die während des Kurses gesammelt werden (‚Anwesenheit bei Abschluss des ersten Studienteils') und die Ergebnisse der drei Tests miteinander verglichen, die in den Clusteranalysen die Grundlage für die Analysen bilden. Für die Teilnehmenden innerhalb der Teilstichprobe werden darüber hinaus weitere Hintergrundinformationen berichtet, etwa Sprachkenntnisse oder die in schulischen, beruflichen oder universitären Ausbildungsverhältnissen verbrachten Jahre. Die 44 Teilnehmenden waren im Schnitt zu Beginn des Unterrichts 24,39 Jahre alt (*Mdn* = 24; Modus = 24; SD = 4,682; n=44). Abbildung 3 zeigt die Altersverteilung der Teilstichprobe:

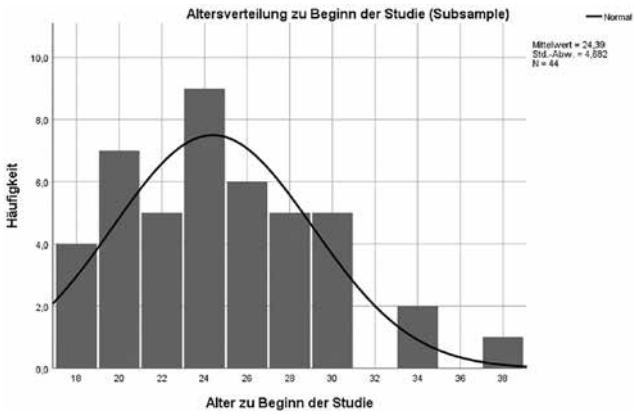


Abbildung 3: Altersverteilung zu Beginn der Studie (Teilstichprobe, n=44)

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Von den 44 Teilnehmenden gaben 38 ihr Geschlecht als männlich und 6 ihr Geschlecht als weiblich an. 24 der 44 Teilnehmenden besuchten Syntaxkurse und 20 Semantikkurse (vgl. Tabelle 6).

Gruppeneinteilung in Bezug auf experimentelle Manipulation (Teilstichprobe)					
		Häufigkeit	**Prozent**	**Gültige Prozent**	**Kumulative Prozent**
Gültig	Syntax	24	54,5	54,5	54,5
	Semantik	20	45,5	45,5	100,0
	Gesamt	44	100,0	100,0	

Tabelle 6: Gruppeneinteilung in Bezug auf die experimentelle Manipulation innerhalb der Studie. Gezeigt wird die Anzahl der Teilnehmenden in den Syntax- und Semantikgruppen innerhalb der Teilstichprobe (n=44).

Die Teilnehmenden der Teilstichprobe besuchten im hier beobachteten Zeitraum im Schnitt 485,36 Stunden Unterricht, was einer Anwesenheitsquote von 81,9 % entspricht (*Mdn* = 492; SD = 80,717).

Die Lernenden innerhalb der Teilstichprobe gaben in der Selbstauskunft folgende Sprachkenntnisse (L2) an: Englisch (n=27) und Kurdisch (n=2). 13 Personen nannten keine Fremdsprachenkenntnisse, für drei sind keine Daten vorhanden. Von den 29 Personen, die Fremdsprachenkenntnisse angaben, nannten sechs Französischkenntnisse und fünf Türkischkenntnisse als weitere Sprachkenntnisse. Eine Person gab Italienischkenntnisse an. Anhand der Selbstauskünfte und in Verbindung mit einem Englisch-Test wurden 15 Teilnehmende als fortgeschritten und 28 als nicht fortgeschritten in einer anderen als der L1 klassifiziert, bei zwei konnte keine Aussage getroffen werden, da beide Datenpunkte fehlen.[153] Im Schnitt wurden 15,64 Jahre in Ausbildungsverhältnissen verbracht (Kindergarten, Schule, Universität, berufliche Weiterbildung), bevor der Deutschkurs gestartet wurde (*Mdn* = 16; Mode = 16, SD = 3,696; n=39). Die minimale Zahl an Jahren in Ausbildungsverhältnissen, die von den Teilnehmenden angegeben wurde, ist neun, die maximale 28. Für fünf Teilnehmende liegen keine entsprechenden Daten vor. 33 Personen gaben darüber hinaus an, die Universität besucht oder eine berufliche Ausbildung begonnen zu haben. Für elf Teilnehmende sind hier keine Daten vorhanden bzw. es wurde angegeben, dass noch keine akademische Bildungserfahrung vorhanden sei.

153 Hier wurden die Ergebnisse der Englisch-Tests mit den Selbstaussagen aus den Fragebögen zum Sprachhintergrund in Verbindung gesetzt, da nicht für alle Teilnehmende Ergebnisse der Englisch-Tests vorliegen. Für eine Bewertung als ‚fortgeschritten' musste in den Englisch-Sprachtests mindestens das Niveau B1 erreicht werden. Nur wenn dieses Niveau erreicht wurde und die eigenen Fähigkeiten auf einer 10-Punkt Skala mindestens mit dem Wert 8 eingeschätzt wurden, kam es zu einer Klassifizierung der Teilnehmende als fortgeschritten.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

32 der Teilnehmenden (ca. 75 % der gesamten Stichprobe) gaben als Heimatland Syrien an, vier Irak, drei Libanon und eine Person Jordanien. Außerdem wurde dreimal keine Staatsangehörigkeit angegeben.

Es fanden sich Teilnehmende aus allen sechs Kursen unter den 44 Teilnehmenden der Teilstichprobe. Auffällig ist, dass die Verteilung der Teilnehmenden über die sechs Kurse nicht ganz gleichmäßig ist. Da alle Kurse nach demselben Curriculum unterrichtet wurden, sind Gruppenunterschiede zwischen den einzelnen Kursgruppen unwahrscheinlich, aber nicht ausgeschlossen. Diese potenziellen Gruppenunterschiede wurden an dieser Stelle nicht weiter untersucht. Analysen hinsichtlich der Gruppen finden sich im Kapitel zu den Ergebnissen der Clusteranalyse (vgl. Kapitel 4.1.3). Siehe hierzu Tabelle 7:

Teilnehmende der Teilstichprobe in Studienphase 1					
		Häufigkeit	Prozent	Gültige Prozent	Kumulative Prozent
Gültig	Kurs 1 (Syntax)	5	11,4	11,4	11,4
	Kurs 2 (Syntax)	11	25,0	25,0	36,4
	Kurs 3 (Semantik)	8	18,2	18,2	54,5
	Kurs 4 (Semantik)	6	13,6	13,6	68,2
	Kurs 5 (Syntax)	8	18,2	18,2	86,4
	Kurs 6 (Semantik)	6	13,6	13,6	100,0
	Gesamt	44	100,0	100,0	

Tabelle 7: Teilnehmende innerhalb der einzelnen Kursgruppen. Gezeigt wird die Aufteilung in Bezug auf die Teilstichprobe (n=44).

3.2.2 Unterschiede zwischen Teilstichprobe und ausgeschlossenem Datenset

Die 21 Teilnehmenden, deren Datensets aus diesen Analysen ausgeschlossen wurden, waren im Schnitt 24,00 Jahre alt (*Mdn* = 22; Modus 20; SD = 4,919; n=21). Von diesen Teilnehmenden gaben 16 (76,2 %) ihr Geschlecht als männlich und 5 (23,8 %) ihr Geschlecht als weiblich an. Acht der ausgeschlossenen Teilnehmenden besuchten Syntaxkurse und 13 Semantikkurse. Diese Teilnehmenden besuchten im Erhebungszeitraum im Schnitt 431,21 Stunden Unterricht, was einer Anwesenheitsquote von 70,5 % entspricht (*Mdn* = 441; SD = 76,70; n=19, zwei Teilnehmende wurden für die Berechnung der Anwesenheit nicht berücksichtigt, da sie den Kurs erst nach einigen Wochen besuchten) (vgl. Tabelle 8).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

	Ausgeschlossene Teilnehmende (n=21) (für Anwesenheit n= 19)			Teilstichprobe (n=44)		
	Mittelwert	Median	Standard-abwei-chung	Mittelwert	Median	Standard-abwei-chung
Alter	24	22	4,919	24,39	24	4,682
Experimentelle Manipulation	0,62	1	0,498	0,45	0	0,504
Geschlecht	0,24	0	0,436	0,20	0	0,408
Anwesenheit (B1)	431,21	441,00	76,702	485,36	492,00	80,717

Tabelle 8: Vergleich zwischen ausgeschlossenen Teilnehmenden und Teilstichprobe (B1)

- Für den Vergleich zwischen der Teilstichprobe und den ausgeschlossenen Teilnehmenden in Bezug auf die Variable ‚Alter' mit einem Unterschied von 0,386 BCa 95 % KI [−2,742, 2,047] ist der t-test nicht signifikant $t(63) = -0{,}301$, $p = 0{,}765$. Die Effektgröße d beträgt −0,082 und stellt damit eine sehr kleine Effektgröße dar.
- In Bezug auf die Variable ‚Experimentelle Manipulation' ergibt sich der Unterschied 0,165 BCa 95 % KI [−0,096; 0,434]. Für den t-test liefert dieser Unterschied kein signifikantes Ergebnis $t(63) = 1{,}241$, $p = 0{,}222$. Es ergibt sich mit $d = 0{,}339$ eine kleine bis mittlere Effektgröße.
- Für die Variable ‚Geschlecht' mit einem Unterschied von 0,034 BCa 95 % KI [−0,188; 0,255] zwischen den beiden Gruppen ergibt sich kein signifikanter Unterschied $t(63) = 0{,}296$, $p = 0{,}769$. Es ergibt sich eine Effektgröße von $d = 0{,}096$.
- Deutliche Unterschiede ergeben sich für die Variable ‚Anwesenheit', die die besuchten Unterrichtsstunden bis zum Goethe-Zertifikat B1 codiert. Die Teilstichprobe besuchte im Schnitt über 50 Unterrichtseinheiten mehr als die ausgeschlossenen Teilnehmenden. Der Unterschied von −54,153 BCa 95 % KI [−95,043; −12,142] ist signifikant $t(61) = -2{,}480$, $p = 0{,}016$. Hier ergibt sich ein mittlerer bis großer Effekt mit einer Effektstärke von −0,681.

Für drei der vier an dieser Stelle untersuchten Variablen ergeben sich zwischen den ausgeschlossenen Teilnehmenden und den Teilnehmenden innerhalb der Teilstichprobe keine signifikanten Unterschiede. Für die Variable ‚Anwesenheit' ergibt sich ein signifikanter Unterschied mit einem mittleren Effekt. Die Teilnehmenden der Teilstichprobe besuchten deutlich mehr Unterricht, was sich durchaus auf ihre Leistungen und auf die später untersuchten syntaktischen Variablen auswirken könnte. Insgesamt könnte die Situation vorliegen, dass in der Teilstichprobe eine Gruppe von Lernenden gebildet wurde, die im Vergleich zur Gesamtgruppe aus stärkeren Lernenden bzw. aus Lernenden, die mehr Unterricht besucht haben, besteht. Wie bei den Abbrüchen handelt es sich sicherlich beim Besuch des Unterrichts ebenfalls um eine multikausal beeinflusste Variable. Es können in der vorliegenden Studie keine Gründe für den Besuch

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

oder den Nicht-Besuch des Unterrichts geklärt werden. Der Zusammenhang zwischen Unterrichtsbesuch und Lernerfolg steht zudem nicht im Zentrum der vorliegenden Untersuchung.

3.2.3 Datengrundlage für Forschungsfragen 1, 2 und 3: Sprachtests

Zur Beantwortung der Forschungsfragen 1, 2 und 3 wurden zunächst Daten zu drei Zeitpunkten herangezogen, die sich zum einen auf die Sprachniveaus der Teilnehmenden bezogen und dabei an GER-Level angepasst waren und zum anderen an der Lernzeit der Teilnehmenden in Stunden orientiert waren. Es wurden schriftliche Daten aus drei Tests analysiert: Ein Test, der zwischen dem A1- und A2-Niveau nach etwa 200 Unterrichtsstunden abgelegt wurde (im Folgenden A1/A2), ein A2-Test, der nach etwa 400 Unterrichtsstunden durchgeführt wurde und die Prüfung Goethe-Zertifikat B1 Deutsch des Goethe-Instituts, die nach etwa 600 Unterrichtsstunden absolviert wurde (vgl. Tabelle 9).[154]

Zeitpunkt	Durchgeführter Sprachtest	Lernzeit in Unterrichtseinheiten
TP1	A1/A2-Test	nach etwa 200 UE
TP2	A2-Test	nach etwa 400 UE
TP3	Goethe-Zertifikat B1	nach etwa 600 UE

Tabelle 9: Übersicht über die durchgeführten Sprachtests in der 1. Studienphase (schematisch)

Die ersten beiden Tests wurden im Rahmen der Studie erstellt und nur zu einem einzelnen Zeitpunkt durchgeführt.[155] Das Goethe-Zertifikat B1 hingegen ist eine

154 Tabelle 9 zeigt den schematischen Ablauf. Da die Lernenden in sechs verschiedenen Kursgruppen unterrichtet wurden, die jeweils zu unterschiedlichen Zeitpunkten starteten (vgl. Kapitel 3.1.3), unterscheiden sich sowohl die Zeitpunkte als auch die tatsächlich zu diesen Zeitpunkten besuchten UE. Im Folgenden werden die Durchschnittswerte für die drei Zeitpunkte bezogen auf die ausgewählte Stichprobe (n=44) aufgeführt. Zeitpunkt 1 ist für die Lernenden der Teilstichprobe im Schnitt nach 15,23 Wochen (*Mdn* = 16; SD = 1,236). Die Lernenden besuchten bis zu diesem Zeitpunkt im Schnitt 244,52 Unterrichtseinheiten (*Mdn* = 259,5; SD = 55,031). Zeitpunkt 2 ist für die Teilnehmenden im Schnitt nach 25,14 Wochen (*Mdn* = 25; SD 1,464) und 390,00 besuchten UE (*Mdn* = 398,5; SD = 87,285). Zeitpunkt 3 ist nach durchschnittlich 29,43 Wochen (*Mdn* = 29; SD = 1,13389) und 485,36 besuchten UE (*Mdn* = 492; SD = 80,717). Zwischen Zeitpunkt 1 und Zeitpunkt 2 liegen etwa neun Wochen und zwischen Zeitpunkt 2 und 3 knapp vier Wochen. Der zeitliche Abstand zwischen Zeitpunkt 1 und Zeitpunkt 3 beträgt ca. 13 Wochen.

155 Für die Erstellung dieser beiden Tests sei Robin Ide ganz herzlich gedankt. Aufgrund großer zeitlicher Limitationen konnten die beiden Tests vor ihrem Einsatz in der Studie nicht pilotiert bzw. validiert werden.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

offizielle und standardisierte Sprachprüfung, die zusätzlich durch eine Probeprüfung vorbereitet und entlastet wurde.[156] Aus Gründen des Datenschutzes bzw. weil die verwendeten Testformate zum Zeitpunkt der Studie noch in regelmäßigem Gebrauch waren, konnte das Goethe-Institut nicht die tatsächlichen Testaufgaben und die von den Lernenden in der Prüfung produzierten Texte zur Verfügung stellen. Deshalb wurden in den folgenden Auswertungen die Textproduktionen aus den Probeklausuren verwendet, die zwei Tage früher durchgeführt wurden (vgl. zu möglichen Einschränkungen durch die Auswahl dieser Texte die Kapitel 3.1.6.2 und 5). Es ist durchaus anzunehmen, dass die Motivation in der tatsächlichen B1-Prüfung höher war als in den Probeklausuren, was deren Aussagekraft einschränken könnte. Für die Texte, die zur Probeprüfung von den Lernenden produziert wurden, ist zusätzlich anzumerken, dass nicht alle Teilnehmenden dieselben Aufgabenstellungen bearbeiteten, da die Prüfungen von den ersten vier Gruppen und von den letzten beiden Gruppen mit großem zeitlichem Abstand (Dezember 2016 und Februar 2017) (vgl. Tabelle 4 in Kapitel 3.1.3) durchgeführt wurden. Die verwendeten Aufgabenstellungen sind unten aufgeführt.

Alle Testitems zur schriftlichen Produktion sind Prüfungsformaten entnommen oder an solche angelehnt, die einerseits am GER orientiert und einem „handlungsorientierten Lehr- und Lernansatz verpflichtet" (Glaboniat / Perlmann-Balme / Studer 2013:11) sind und andererseits allgemeine kommunikative Kompetenz testen, wobei es sich nur bei der Prüfung zum Goethe-Zertifikat B1 um eine standardisierte Sprachprüfung handelt. Ziel solcher schriftlichen Aufgaben ist die Überprüfung „[s]prachliche[r] Handlungsfähigkeit […] anhand von spontan produzierten und nicht auswendig reproduzierten Performanzen […]" (Hinger / Stadtler 2018: 89). Im Modul Schreiben des Goethe-Zertifikats B1 werden diese lernersprachlichen Performanzen durch „Textsorten und Kontexte […], die der Zielgruppe im realen Alltag begegnen und ihr vertraut sind, beispielsweise Forumsbeiträge im Internet und kurze Mitteilungen per E-Mail" (Glaboniat et al. 2013: 80; vgl. Europarat 2001: 67–68) realisiert. Die einzelnen Prüfungsmodule werden in Prozentangaben bewertet. Zum Bestehen eines der Module müssen in diesem Modul mindestens 60 % erreicht werden.

Bei allen drei (bzw. vier) Aufgaben handelt es sich nach Hinger / Stadler (2018) um kontextualisierte Schreibaufgaben, die die Aufgabenstellung in einen bestimmten kommunikativen Kontext einbetten. Die Aufgabenstellungen inkorporieren in der Regel spezifische sprachliche Operatoren, die bei der Um-

156 Für die Teilnehmenden ist hier von einer *high stakes*-Prüfung auszugehen, da das Goethe-Zertifikat B1 die erste Sprachprüfung darstellte, die offiziell ihr Sprachniveau zertifizierte. Im Falle eines erfolgreichen B1-Abschlusses musste darüber hinaus kein Integrationskurs mehr besucht werden. Viele der Teilnehmenden verknüpften (oft berechtigte) Hoffnungen auf Jobs und erste Schritte in Richtung Ausbildung mit dem Erreichen des Sprachniveaus B1.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

setzung der Aufgabe erwartet werden und daher die möglichen Antworten vorstrukturieren (vgl. Hinger/Stadler 2018: 94). Das bedeutet, dass die Texte eher ‚erfasst' als ‚erhoben' wurden. Schramm definiert ‚Erfassen' als „das Gewinnen von LernerInnentexten, die unabhängig von einer Studie im (Schul-) Alltag ohnehin entstehen" (Schramm 2018: 35). ‚Erheben' hingegen bezieht sich auf Fälle, „in denen die Texte aufgrund des Eingriffs der Forschenden von DaZ-Schreibenden produziert werden" (ebd.). Schramm macht weiterhin darauf aufmerksam, dass „[s]owohl beim Erheben als auch beim Erfassen [...] dem Kontext bzw. der Aufgabenstellung eine wichtige Rolle für das Schreibhandeln zu[kommt]" (ebd.).

In den vorliegenden Aufgaben wurde von den Lernenden in zwei Fällen erwartet, eine E-Mail an einen Freund bzw. eine Freundin zu schreiben und von eigenen Erfahrungen zu berichten (A1/A2 und B1; vgl. Abbildung 5, Abbildung 7 und Abbildung 8), wobei im ersten Fall eine Einladung nach Leipzig formuliert und im zweiten Fall die Mitgliedschaft in einem Fitnessstudio geschildert werden sollte. Im dritten Test (A2; vgl. Abbildung 6) bestand die Aufgabe darin, einen Blog-Beitrag über einen Besuch in Berlin zu schreiben. Während im A1/A2-Test und beim Goethe-Zertifikat B1 die Operatoren in Fragen eingebettet sind, tauchen sie im A2-Test in Form von Arbeitsaufträgen auf.

Da zusätzlich aus dem oben genannten Grund die Textproduktionen im Kontext des Goethe-Zertifikats B1 nicht direkt mit den Textproduktionen der beiden anderen Tests verglichen werden konnten, wurden die Ergebnisse der Teiltests Schreiben der B1-Probeprüfung und des Goethe-Zertifikats B1 statistisch miteinander verglichen. Um zu überprüfen, wie vergleichbar die Ergebnisse der Probeprüfung mit der tatsächlichen Prüfung waren, wurde nach einer grafischen Überprüfung (vgl. Abbildung 4) für die Teilstichprobe von 43 Teilnehmenden in Bezug auf die Schreibteile aus der Probeprüfung und aus der Prüfung zum Goethe-Zertifikat B1 zusätzlich eine Korrelationsanalyse durchgeführt.[157] Eine teilnehmende Person wird ausgeschlossen, da sie zwar die Probeprüfung, aber nicht die Prüfung zum Goethe-Zertifikat B1 mitabgelegt hatte.

Die Korrelationsanalyse zeigt eine signifikante positive Korrelation zwischen den Ergebnissen der beiden Schreibteile mit großer Effektstärke, $r = 0{,}616$ BCa

157 Hier ist zu beachten, dass die Probeprüfung von Lehrkräften korrigiert wurde, die innerhalb der Studie angestellt waren. Alle diese Lehrkräfte hatten entsprechende Prüfungsfortbildungen des Goethe-Instituts besucht. Die Testbögen des Goethe-Zertifikats B1 wurden von internen Prüferinnen und Prüfern des Goethe-Instituts begutachtet und bewertet. Wenn davon ausgegangen wird, dass sich die sprachliche Leistung von Lernenden innerhalb von zwei Tagen, in denen kein Unterricht mehr stattfand, nicht entscheidend verändert, dann gibt die errechnete Korrelation Hinweise auf ähnliche Bewertungen der Teilnehmenden durch die unterschiedlichen Prüfer/-innen.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

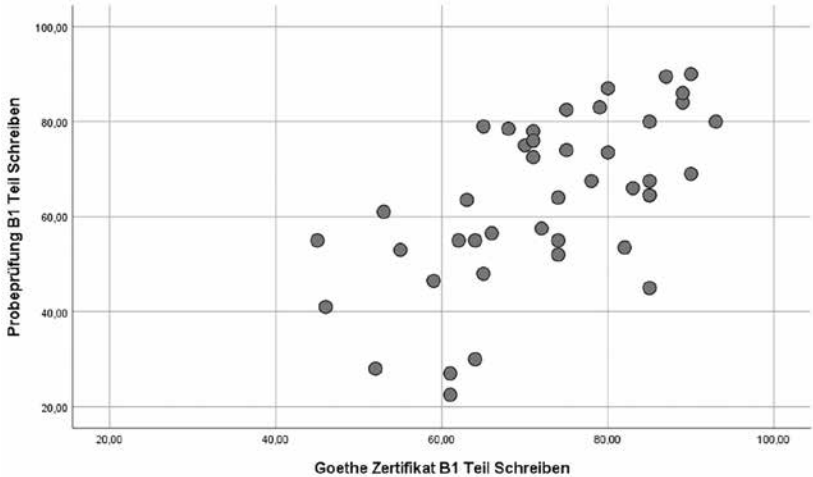


Abbildung 4: Streudiagramm zum Vergleich der schriftlichen Ergebnisse in der Probeprüfung und in der Prüfung zum Goethe-Zertifikat B1; die Ergebnisse werden in Prozentangaben dargestellt.

95 % KI [0,447; 0,751][158], $p = 0{,}000$.[159] Prinzipiell kann davon ausgegangen werden, dass Teilnehmende, die höhere Ergebnisse in der einen Prüfung erreichten, ebenso hohe Ergebnisse in der anderen Prüfung erreichten, selbst wenn es sich bei der gefundenen Korrelation um eine eher moderate Korrelation handelt.

Die Mittelwerte im Probetest liegen für die Teilstichprobe deutlich und statistisch signifikant unter den Werten in der tatsächlichen Prüfung (vgl. Tabelle 10).[160]

158 Zu den Bias-korrigierten mit Bootstrap ermittelten Konfidenzintervalle vgl. Fußnote 151.

159 Wenn nicht nur die Teilstichprobe, sondern alle verfügbaren Teilnehmendendaten zu diesem Zeitpunkt (N=51) miteinander verglichen werden, ist die Korrelation etwas höher. Es ergibt sich eine signifikante Korrelation zwischen den Ergebnissen der Schreibteile im Probetest und im Goethe-Zertifikat B1 mit großer Effektstärke, $r = 0{,}713$ [0,534; 0,822], $p = 0{,}00$.

160 Ein t-test bei verbundenen Stichproben wurde durchgeführt, um die Unterschiede zwischen den beiden Tests zu untersuchen. Dieser ergibt bei einer Differenz der Mittelwerte von 8,942 zwischen den Ergebnissen des Schreibteils des Goethe-Zertifikats B1 und des Probetests einen signifikanten Unterschied mit $t(42) = -4{,}209$, $p = 0{,}000$. Gleichzeitig ergibt sich ein mittlerer Effekt mit einer Effektstärke $d = -0{,}587$.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Deskriptive Statistiken			
	Mittelwert	Std.-Abweichung	N
Goethe-Zertifikat B1: Teil Schreiben	72,58	12,46	43
Probeprüfung B1: Teil Schreiben	63,30	17,51	44

Tabelle 10: Überblick über die Ergebnisse der schriftlichen Teile der Probeprüfung und des Goethe-Zertifikat B1. Gezeigt werden der Mittelwert und die Standardabweichung in Bezug auf die Teilstichprobe (n=44).

Die Differenz zwischen den beiden Prüfungen ist unter Umständen hauptsächlich auf einige Ausreißer zurückzuführen.[161] Dies wird im Streudiagramm (vgl. Abbildung 4, im Diagramm blau eingekreist) durch die Gruppe von vier Teilnehmenden (TN005, TN043, TN057 und TN00X) verdeutlicht, die in der Probeprüfung deutlich schwächer als in der Prüfung zum Goethe-Zertifikat B1 zwei Tage später abgeschlossen haben. Für diese vier Teilnehmenden lässt sich die vorliegende Differenz weitgehend erklären: TN005 hat die Probeprüfung mitgeschrieben, aber aufgrund seines schlechten Abschneidens die Möglichkeit in Anspruch genommen, die Prüfung zum Goethe-Zertifikat B1 erst etwas später zusammen mit den Gruppen 5 und 6 abzulegen. Die entsprechende Prüfung fand nicht im Dezember, sondern erst im Februar statt. In der Zwischenzeit bereitete sich TN005 mit den anderen Gruppen weiter auf die Prüfung vor.[162] So erklärt sich die niedrigere Teilnehmendenzahl von n=43 in der Prüfung zum Goethe-Zertifikat B1. Die Teilnehmenden TN00X, TN043, TN057, die darüber hinaus als Ausreißer identifiziert wurden, haben jeweils in der Probeprüfung nicht alle Aufgaben des Schreibteils beantwortet, was einer der Gründe für eine deutlich niedrigere Bewertung dieses Prüfungsteils in der Probeprüfung ist.[163] Letztlich ist die hohe Korrelation der beiden Testteile aus Probeprüfung und Goethe-Zertifikat B1 ausschlaggebend für die Entscheidung, alle Textproduktionen der Probeprüfung (n=44) in die Analyse mitaufzunehmen.

161 Als Ausreißer (*outlier*) werden Datenpunkte innerhalb einer Datenreihe bezeichnet, die sich in ihren Eigenschaften grundlegend von den anderen Datenpunkten in einer Datenreihe unterscheiden. Ausreißer können deskriptive Statistiken wie den Mittelwert deutlich verzerren. Außerdem beeinflussen und verzerren sie inferenzstatistische Verfahren, die anhand der betreffenden Datenreihen durchgeführt werden (vgl. Field 2018: 227–229).

162 Ein solches Vorgehen ist zwar nicht optimal, was die Vergleichbarkeit der Ergebnisse angeht, erschien der Forschungsgruppe aber vor dem Hintergrund einer letztlich erfolgreich bestandenen Prüfung für die Teilnehmenden vertretbar.

163 Ein Ausschluss dieser vier Teilnehmenden würde die Differenz der Mittelwerte von ~8,94 auf ~6,51 reduzieren. Da die Teilnehmenden nicht aus den weiteren Analysen ausgeschlossen wurden, wurden an dieser Stelle keine weiteren Tests durchgeführt.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

A1/A2-Test

Schreiben – Teil 2

Schreiben Sie an einen Freund/ eine Freundin eine E-Mail.

- Sie laden Ihren Freund/ Ihre Freundin nach Leipzig ein.
- Welche Aktivitäten planen Sie?
- Was wollen Sie sich ansehen?

Schreiben Sie zu jedem Punkt zwei Sätze!

Abbildung 5: Schreibaufgabe aus dem A1/A2-Test; im Rahmen der Studie erstellt.

A2-Test

Schreiben – Teil 2

Sie **waren** letztes Wochenende in Berlin und schreiben in Ihrem Internetblog darüber. Erzählen Sie:

- Was haben Sie in Berlin gemacht? (Aktivitäten, Sehenswürdigkeiten?)
- Wie war das Wetter in Berlin?
- Was hat Ihnen besonders gefallen?
- Finden Sie auch einen passenden Titel für Ihren Blog!

Schreiben Sie zu jedem Punkt zwei Sätze!

Abbildung 6: Schreibaufgabe aus dem A2-Test; im Rahmen der Studie erstellt.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Prüfung Goethe-Zertifikat B1 a)[164]

Die beiden dem Goethe-Zertifikat B1 entnommenen Schreibaufgaben a und b (siehe unten) sind jeweils die erste von drei Aufgaben aus dem Modul Schreiben des Goethe-Zertifikats B1.[165] In diesen Aufgaben geht es darum, dass die Prüfungsteilnehmenden ‚interagieren' (hier und im Folgenden: Glaboniat et al. 2013: 82), indem sie eine persönliche Mitteilung zur Kontaktpflege verfassen. Die Aufgaben orientieren sich an folgender Kann-Beschreibung zur „Schriftliche[n] Interaktion allgemein" (Europarat 2001:86) aus dem GER:

> Kann in persönlichen Briefen und Mitteilungen einfache Informationen von unmittelbarer Bedeutung geben oder erfragen und dabei deutlich machen, was er/sie für wichtig hält (ebd.).

Der geforderte Texttyp ist eine informelle und private E-Mail. Die Teilnehmenden werden dazu aufgefordert, einen freien Text mit ca. 80 Wörtern zu schreiben, in dem sie ihre persönlichen Erlebnisse beschreiben, begründen und zuletzt dem Rezipienten oder der Rezipientin darauf aufbauend einen Vorschlag unterbreiten (etwa für eine gemeinsame Unternehmung; vgl. Glaboniat et al. 2013: 83). Als Zeitrahmen sind für diese Aufgabe innerhalb der Prüfung 20 Minuten vorgesehen. Da dies die erste von drei Aufgaben ist, ist unter Umständen davon auszugehen, dass Teilnehmende mehr Zeit für diese Aufgabe aufwenden. In der Aufgabenstellung wird der Schreibanlass situiert und mit drei Handlungsanweisungen versehen (vgl. ebd.). In der ersten Aufgabe, die von den Kursgruppen 1–4 bearbeitet wurde, ging es darum, über die eigene Mitgliedschaft in einem Fitnessstudio zu berichten und eine Freundin oder einen Freund davon zu überzeugen, sie zu begleiten.

Schreiben

Aufgabe 1 Arbeitszeit: 20 Minuten

Sie sind seit zwei Wochen Mitglied in einem Fitnessstudio und berichten einem Freund/einer Freundin von Ihren Erfahrungen.

- Beschreiben Sie, wie es Ihnen gefällt.
- Begründen Sie: Was machen Sie besonders gern und warum?
- Schlagen Sie Ihrem Freund/Ihrer Freundin vor, einmal mitzukommen.

Schreiben Sie eine E-Mail (circa. 80 Wörter).
Schreiben Sie etwas zu allen drei Punkten.
Achten Sie auf den Textaufbau (Anrede, Einleitung, Reihenfolge der Inhaltspunkte, Schluss).

Abbildung 7: Schreibaufgabe a aus Modellsatz für das Goethe-Zertifikat B1 a

164 Der entsprechende Modellsatz ist nicht mehr frei verfügbar und es konnte kein Nachdruck der Aufgabe ermittelt werden.

165 Für eine genaue Beschreibung des Moduls Schreiben des Goethe-Zertifikats B1 vgl. Glaboniat et al. (2013: 80–94).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Prüfung Goethe-Zertifikat B1 b)[166]

In der zweiten Aufgabe, die von den Kursgruppen 5 und 6 bearbeitet wurde, ging es darum, von einem Besuch auf einem Stadtfest zur berichten, zu begründen, was dort schön war und ein Treffen mit einer Freundin bzw. einem Freund zu vereinbaren.

Zertifikat B1a **Schreiben**

Aufgabe 1 Arbeitszeit: 20 Minuten

Sie waren am Wochenende auf einem Stadtfest. Ein Freund/Eine Freundin von Ihnen konnte nicht mitkommen, weil er/sie krank war.

- Beschreiben Sie: Wie war das Stadtfest?
- Begründen Sie: Was hat Ihnen am besten gefallen und warum?
- Machen Sie einen Vorschlag für ein Treffen.

Schreiben Sie eine E-Mail (circa. 80 Wörter).
Schreiben Sie etwas zu allen drei Punkten.
Achten Sie auf den Textaufbau (Anrede, Einleitung, Reihenfolge der Inhaltspunkte, Schluss).

Abbildung 8: Schreibaufgabe b aus Modellsatz für das Goethe-Zertifikat B1.

3.2.4 Methodisches Vorgehen: Emergenz- und Profilanalyse

Im folgenden Abschnitt werden die Analyseschritte hinsichtlich der Emergenz- und Profilanalysen, die anhand der lernersprachlichen Texte dieser Studie durchgeführt wurden, erklärt. Während bei der Aufbereitung und der initialen Analyse der Texte den Vorgaben im Rahmen der Profilanalyse nach Grießhaber gefolgt wurde, wurden im nächsten Analyseschritt zwei unterschiedliche Emergenzkriterien in Bezug auf die Texte angewendet, um herauszuarbeiten, welchen Einfluss Emergenzkriterien auf die Interpretation lernersprachlicher Texte haben könnten.

Ein grundlegendes Prinzip zur Operationalisierung von Erwerb ist die Emergenz einer bestimmten Struktur.[167] Diese bezeichnet das erste systematische Auftreten bzw. die erste systematische und produktive Verwendung einer Struktur (vgl. Pallotti 2007: 366; Pienemann 1998; Clahsen et al. 1983). Sobald diese Struktur in einem (mündlichen) Text systematisch auftritt, kann der Text der entsprechenden Erwerbsstufe zugeordnet werden. Es existieren eine Reihe sehr unterschiedlich operationalisierter Emergenzkriterien, die unterschiedlich motiviert, aber kaum empirisch begründet werden (können). Pienemanns (1998) Operationalisierung von Emergenz setzt voraus, dass eine Struktur mindestens einmal korrekt in einem Text verwendet wurde, sofern gleichzeitig mindestens vier obligatori-

166 Entnommen aus Modellsatz 1, nicht mehr frei verfügbar, jetzt aber abgedruckt in Modellprüfung 1 Goethe-/ÖSD-Zertifikat B1 (2018). Unter: https://www.klett-sprachen.de/modellpruefung–1-goethe-oesd-zertifikat-b1/t–1/NP00810000030 [23.02.2023].

167 Das Konzept der Emergenz ist relevant für verschiedene SLA-Theoriestränge (vgl. Kapitel 2.4 und 2.5).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

sche Kontexte für dieselbe Struktur in diesem Text produziert werden. Grießhaber (2012) operationalisiert Emergenz als die dreimalige korrekte Verwendung einer Struktur innerhalb eines Textes. Beispielsweise ordnet Grießhaber beim dreimaligen Vorkommen von finiten Verben in Zweitstellung einen Text der Profilstufe 1 zu (vgl. Tabelle 2 in Kapitel 2.4.3 zu den unterschiedlichen Erwerbs- bzw. Profilstufen der *Processability Theory* und der Profilanalyse). Um die Emergenz von syntaktischen Wortstellungsmustern nachweisen zu können, werden in einem ersten Schritt die Texte in kleinere Einheiten segmentiert, denen dann jeweils Erwerbsstufen zugeordnet werden. In den Ansätzen der *Processability Theory* hat sich in Bezug auf die Analyse des Deutschen als L2 die Einteilung in minimale satzwertige Einheiten durchgesetzt, da in diesen alle verbalen Wortstellungsmuster beobachtet werden können (vgl. Jansen 2008). Dieses Vorgehen übernimmt Grießhaber bei der Operationalisierung seiner Form der Profilanalyse. Grießhaber (2005) macht darauf aufmerksam, dass die Profilanalyse interpretative Züge aufweise und teilweise Entscheidungen bei der Einstufung eines Texts auf einer niedrigeren Stufe getroffen werden könnten, obwohl (ausreichend) Belege für die Emergenz von Strukturen auf einer höheren Stufe gefunden würden (vgl. Grießhaber 2005: 16). Dies kann vor allem in didaktischen Kontexten und, falls Konsequenzen (wie etwa ein bestimmtes Betreuungsangebot) für Lerner/-innen von der jeweiligen Einstufung abhängen, ein sinnvolles Vorgehen sein. In der vorliegenden Studie wurde versucht, das interpretative Element der Profilanalyse soweit wie möglich auszuschließen. Entscheidungen über Profilstufenzuordnungen wurden rein quantitativ getroffen.

In einem ersten Arbeitsschritt zur Aufbereitung der Texte wurden zunächst alle von den Lernenden in den oben beschriebenen Testsituationen produzierten Texte transkribiert und in minimale satzwertige Einheiten (MSE) segmentiert (vgl. Grießhaber 2016: 154; vgl. Abbildung 9).[168]

[1] Gestern ich habe zum Zoo gegangen	(3/0)	2
[2] und Ich hatte viel Spaß,	(1/1)	1
[3] ich habe viele Tiere gesehen.	(2/2)	2
[4] Ich habe um 9:00 Uhr gegangen mit meine Freunden und miT meine lehrerin	(2/2)	2
[5] und meine lehrerin haTTesT ein Plan Für zoo	(1/1)	1
[6] Der zoo ist viele Groß	(1/1)	1
[7] und wir haben ETwa 6 sTunden bis wir FerTisch,	(1/1)	1
[8] ich habe der Tieger gesehen	(2/2)	2
[9] und das ich gesTern foTografierT habe	(4/0)	0
[10] Es, isT sehr süß	(1/1)	1
[11] und Ich habe Der ElefanT gesehen	(2/2)	2
[12] und wir habe viele Pause gemachT	(2/2)	2
[13] und am Ende (ich mag meine lehrerin Tereza und SchTifani) für die Reise	(3/0)	1

Abbildung 9: Beispiel für Transkription und Segmentierung in minimale satzwertige Einheiten (TN085: Mein Besuch im Zoo).

168 Hier existieren in der deutschsprachigen Forschung weitere Beschreibungen. Petersen (2014) spricht in diesem Zusammenhang von „Teilsätzen" (Petersen 2014: 80).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Grießhaber (2016) betrachtet minimale satzwertige Einheiten als „Kern aus finitem Verb und Subjekt" (Grießhaber 2016: 154). Die hier operationalisierte Definition folgt damit der vor allem in englischsprachigen Kontexten verwendeten *clause*-Definition (vgl. Vyatkina 2013: 18; vgl. Lu 2011). In den Variablenbezeichnungen ist dementsprechend das Kürzel *C* für *Clause* zu finden (z. B. MLC = *Mean Length of Clause*). Damit geht die Profilanalyse wie die PT vor, auf der sie maßgeblich aufbaut (vgl. Pienemann 1998; Jansen 2008).[169] Jeder identifizierten minimalen satzwertigen Einheit wurde im nächsten Schritt eine syntaktische Profilstufe zugeordnet (vgl. Kapitel 2.4.3 zur Profilanalyse). Erfüllte eine MSE die Anforderungen an eine bestimmte Stufe nicht, wurde im nächsten Schritt die direkt darunterliegende Stufe geprüft. Das bedeutet, dass MSE, die auf einer Stufe eine Struktur zeigen, die der Stufe noch nicht entspricht, auf einer darunterliegenden Stufe eingeordnet werden können und MSE mit teilweise abweichenden Strukturen nicht automatisch als fehlerhaft gekennzeichnet und aus der Analyse ausgeschlossen werden. Die ermittelten Stufen für die einzelnen MSE verdeutlichen zusammengenommen das syntaktische Profil der untersuchten Texte. Aufgrund der so ermittelten syntaktischen Profile der Texte wurden dann die syntaktischen Niveaustufen der einzelnen lernersprachlichen Texte bestimmt. Bei der Anwendung der Profilanalyse nach Grießhaber wird davon ausgegangen, dass bestimmte Strukturen als erworben gelten können, wenn diese dreimal in einem mündlichen oder schriftlichen Text verwendet werden. Jedem Text wird die höchste jeweils als erworben herausgearbeitete syntaktische Profilstufe zugeordnet. Wenn in einem Text viermal eine Inversionsstruktur (Profilstufe 3) auftaucht, aber nur zwei (und nicht drei) Strukturen mit einer Verbendstellung (Profilstufe 4), dann wird dem gesamten Text die syntaktische Profilstufe 3 (Inversion) zugeordnet.

Grießhaber (2012: 176)		
mündlich, schriftlich DaF/DaZ		
6	Integration	–
5	Insertion	–
4	Nebensatz-Endstellung	*Als ich im fitnessstudio angemeldet* ***habe***, […]. (TN031)
3	Inversion	*Seit zwei Wochen habe ich in einem Fitnessstudio angemeldet.* (TN031)
2	Separation	*Ich* ***muss*** *immer gesund* ***essen****.* (TN031)
1	Finitum	***Meine Muskeln sind*** *jetzt sehr starker.* (TN031)[170]
0	Bruchstücke	*Liebe Grüße* (TN031)

Tabelle 11: Profilstufen nach Grießhaber (2012)

Mithilfe des Friedman-Tests (vgl. Field 2018: 321–328; vgl. Friedman 1937) wurde untersucht, ob sich die durch die Profilanalysen ermittelten Niveaustufen der Teilnehmenden insgesamt über die drei Zeitpunkte hinweg signifikant ändern bzw. unterscheiden. Der Friedman-Test bzw. die Friedman-Varianzanalyse (*Friedman ANOVA: Analysis of Variance*) wird als nicht-parametrischer

[169] Grießhaber (2012: 167) postuliert zwar eine erwerbstheoretische Unabhängigkeit der Profilanalyse, wesentliche Elemente der Annahmen und des Vorgehens wurden jedoch aus der PT übernommen (vgl. Kapitel 2.4.1 und 2.4.3).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Test eingesetzt, der misst, ob verbundene Gruppen sich hinsichtlich der gemessenen Variablen in drei oder mehr Bedingungen unterscheiden (vgl. Field 2018: 321f.). Nicht-parametrische Verfahren werden in der Regel dann eingesetzt, wenn in der untersuchten Stichprobe nicht alle Voraussetzungen für die Anwendung parametrischer Verfahren gegeben sind.[171] Dies war für die hier untersuchte Variable ‚Erreichte Profilstufe' hinsichtlich der Normalverteilung der Fall. In der vorliegenden Studie erlaubte es der Friedman-Test, zu untersuchen, ob sich die Teilnehmenden der Teilstichprobe (n=44, verbundene Gruppe) über die drei Zeitpunkte (=drei Bedingungen) hinsichtlich der jeweils herausgearbeiteten Niveaustufen unterscheiden.[170] [171]

Anschließend an den Friedman-Test wurde anhand der beiden Emergenzkriterien von Pienemann (1998) und Grießhaber (2012) untersucht, wie viele der Teilnehmenden in ihrer Entwicklung über drei Zeitpunkte den vorausgesagten Erwerbssequenzen folgen und ob die beiden Kriterien unterschiedliche Erwerbsverläufe evozieren.[172] Dabei wurde analysiert, ob die einzelnen ‚syntaktischen Erwerbsstufen' innerhalb der Lernertexte in der vorhergesagten Reihen-

170 Kopulasätze wie im Beispielsatz zu Stufe 1 ‚Finitum' in Tabelle 11 bilden für die Bestimmung von Verbstellungsmustern eine durchaus problematische Struktur. In einem Kopulasatz wird „einer Entität eine Eigenschaft (Adjektiv, Adverb,PP, NP) oder eine Lokation (Adverb, PP) zugeschrieben, oder sie wird mit einer anderen Entität (NP) identifiziert" (Czinglar 2014: 159). Für Czinglar (2014) ist außerdem „nicht eindeutig zwischen Subjekt und Prädikat" (ebd.) zu unterscheiden, falls das Prädikat eine Nominalphrase ist. In Czinglars Analysen werden daher einige Kopulakonstruktionen aus den Analysen ausgeschlossen bzw. getrennt von anderen Wortstellungsmustern betrachtet (vgl. ebd.) Grießhaber (2016: 148) geht davon aus, dass Kopula und Kopulakonstruktionen zu Stufe 1 ‚Finitum' zählen, da zwar das logische Prädikat aus zwei Strukturen besteht und eine Klammer bildet, grammatisch jedoch in Kopulakonstruktionen nur ein Verb vorliegt. Grießhaber folgend wurden Kopulakonstruktionen in dieser Arbeit der Stufe 1 ‚Finitum' zugeordnet.

171 Als Voraussetzungen (*assumptions*) für die Anwendung parametrischer Verfahren gelten Linearität, Normalverteilung der Stichprobe, die Varianzhomogenität und statistische Unabhängigkeit (vgl. für eine ausführliche Einführung in die genannten Voraussetzungen und in die Tests zur Überprüfung, ob diese Voraussetzungen in Datenreihen bzw. Stichproben gegeben sind Field 2018: 229–277).

172 Pienemann (1998) setzt als Kriterium für Emergenz voraus, dass eine Struktur einmal korrekt realisiert wird und gleichzeitig mindestens vier obligatorische Kontexte für diese Struktur im untersuchten Text vorliegen. Für Grießhaber (2012) zählt eine Struktur als in einem Text emergent, wenn diese mindestens dreimal produziert wird. Dabei ist nicht ausschlaggebend, in wie vielen obligatorischen Kontexten diese vorliegen. Hier ist zu beachten, dass das Emergenzkriterium nach Pienemann (1998) sich zunächst ausschließlich auf spontan produzierte und gesprochensprachliche Daten bezog, mittlerweile aber in verschiedenen Kontexten ebenfalls auf schriftliche Daten angewendet wird (vgl. hierzu vor allem die Studien in Kapitel 2.4.2).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

folge emergent werden. Bei diesen Analysen wurden jeweils alle produzierten (obligatorischen) Kontexte miteinbezogen und nicht nur die höchste erreichte Profilstufe wie beim Friedman-Test (vgl. zu den obligatorischen Kontexten Kapitel u.a. 2.4.1). Es wurden folglich alle in einem Text erreichten Stufen gleichzeitig betrachtet. Dies ermöglichte, Aussagen darüber zu treffen, ob etwa eine höhere Stufe produziert wurde, während sich für die darunterliegenden Stufen ebenfalls Belege im analysierten Text finden oder ob diese Belege nicht vorhanden sind. War dies der Fall, wurde der betreffende Fall zunächst als ‚nicht den Erwerbsstufen entsprechend' eingestuft. Es ist allerdings möglich, dass eine Struktur in einem Text nicht produziert wurde, obwohl sie von den betreffenden Lernenden beherrscht wurde (vgl. zu dieser grundsätzlichen methodischen Herausforderung die Diskussion in Kapitel 5).

3.2.5 Methodisches Vorgehen: Longitudinale Clusteranalysen

Ausgehend von den Ergebnissen dieser Analysen wurden mit den Daten der 44 Teilnehmenden longitudinale Clusteranalysen (MacIntyre et al. 2017; Pinel / Csizér 2015) durchgeführt, um durch die Klassifizierung in unterschiedliche Gruppen bzw. Cluster bestimmte Entwicklungsprofile herauszuarbeiten (vgl. Baba / Nitta 2014: 30).

Clusteranalysen sind explorative Analysetechniken, die helfen sollen, Subgruppen in einer Stichprobe zu identifizieren (Staples / Biber 2015). Ziel einer Clusteranalyse ist es, hinsichtlich jeweils festzulegender Variablen möglichst ähnliche Gruppen innerhalb einer Stichprobe zu identifizieren und die Varianz innerhalb der unterschiedlichen Gruppen (bzw. Cluster) zu minimieren. In den letzten Jahren gewinnen diese Analysetechniken in der Zweitspracherwerbforschung zunehmend an Bedeutung. Dies verdeutlichen aktuelle Studien aus dem Kontext DaF/DaZ. In der Studie von Cindark / Hünlich 2019 werden die Teilnehmenden von Integrationskursen mittels Clusteranalysen untersucht. Kienberger (2020) und Kienberger et al. (2021) untersuchen anhand von Clusteranalysen den Einsatz verschiedener Strategien zur Erschließung unbekannten Wortschatzes bei DaF-Lernenden an spanischen Universitäten. Murakami (2014) untersucht mit longitudinalen Clusteranalysen individuelle Variation innerhalb der Entwicklung grammatischer Morpheme bei Lernenden des Englischen als L2. Die Studie von Pinel / Csizér 2015 (vgl. Kapitel 2.5.2 zu empirischen Studien innerhalb des CDST-Frameworks) widmet sich ebenfalls mithilfe einer longitudinalen Clusteranalyse den Konstrukten Schreibangst, Selbstwirksamkeit beim Schreiben und Motivation. Es existieren sehr unterschiedliche Clusteranalysen, die sich vor allem hinsichtlich des an die Daten angelegten Proximitätsmaßes und des Gruppierungsverfahrens unterscheiden (vgl. Backhaus et al. 2018: 437). Das Proximitätsmaß misst die Ähnlichkeit bzw. Unähnlichkeit zwischen den Objekten bzw. den Lernenden. Das Gruppierungsverfahren bezeichnet das Verfahren, nach dem eine Zusammenfassung zu Clustern

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

bzw. Gruppen erfolgt (vgl. ebd.). Je nach gewählter Clustermethode unterscheiden sich die durchzuführenden Analyseschritte leicht. In der aus Crowther et al. (2020: 6) entnommenen Abbildung unten werden zwei Verfahren (Hierarchische Clusteranalyse und k-means Clusteranalyse) vergleichend gegenübergestellt (vgl. Abbildung 10).

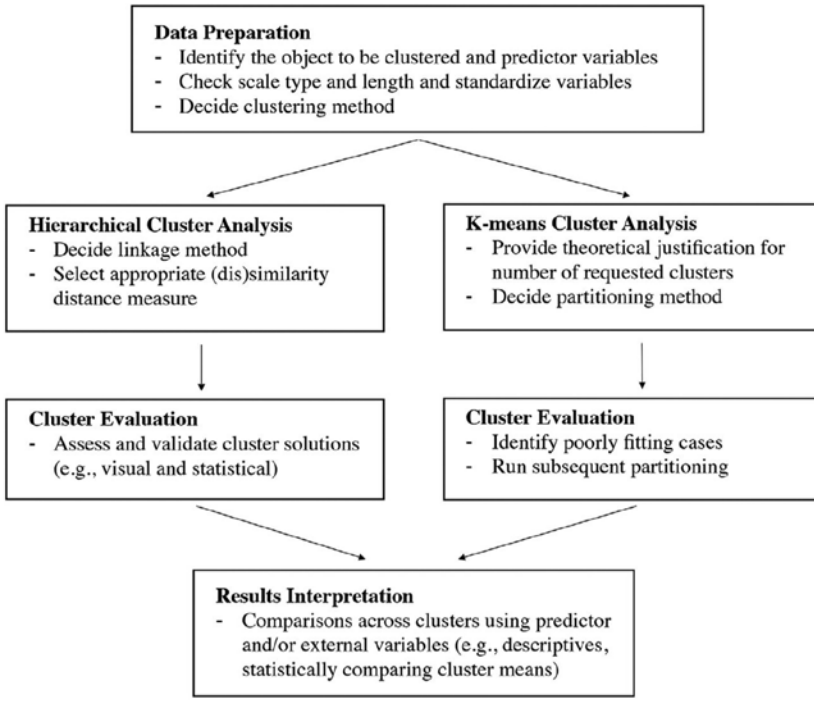


Abbildung 10: Schematischer Ablauf unterschiedlicher Clusteranalysen; Abbildung aus Crowther et al. 2020: 6

In der vorliegenden Analyse wurde eine Kombination aus den beiden oben dargestellten Verfahren (vgl. Abbildung 10) gewählt. Zunächst wurde mithilfe einer hierarchischen Clusteranalyse (HCA) und den Daten von TP1 die optimale Anzahl an Clustern definiert (vgl. für dieses Vorgehen Crowter et al. 2020: 6; vgl. Csizér / Dörnyei 2005; vgl. Papi / Teimouri 2014). Ein Vorteil dieses Verfahrens ist, dass die Struktur der Entwicklung nicht vorgegeben ist und deshalb die aus den Daten erfasste Struktur der tatsächlichen Entwicklung ähnlicher als bei einer vorgegebenen Struktur sein könnte (vgl. Murakami 2014: 136–137). In einem ersten Schritt werden mithilfe eines Proximitätsmaß die Ähnlichkeiten bzw. Distanzen zwischen den Elementen bzw. Variablen, die in die Clusterana-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

lyse eingehen, berechnet. Zur Berechnung wird eine Rohdatenmatrix angelegt, in der alle Lernenden und die jeweiligen Variablenwerte eingetragen werden. Grundsätzlich ist bei der Auswahl der Variablen zu beachten, dass in ihnen eine relativ große Varianz vorhanden sein sollte, um voneinander klar unterscheidbare Cluster bilden zu können. Die Variablen werden während einer Clusteranalyse, falls sie auf unterschiedlichen Skalenniveaus vorliegen, als *z-scores* normalisiert.[173] Als Proximitätsmaß wird die quadrierte euklidische Distanz verwendet, bei der für „jedes Objektpaar die Differenzwerte jeder Eigenschaft quadriert und anschließend addiert" (Backhaus et al. 2018: 450) werden.[174] Bezogen auf Daten von unterschiedlichen Lernenden würde das beispielsweise in einem Sprachtest, bei dem Hören und Lesen getestet wird, heißen, dass zwischen allen Lernenden jeweils die Distanzwerte innerhalb der einzelnen Fertigkeiten berechnet würden. In einem letzten Schritt werden die Elemente (bzw. in der vorliegenden Clusteranalyse die Lernenden) auf der Grundlage der berechneten Ähnlichkeitswerte mit dem Ward-Verfahren fusioniert, wobei alle mithilfe der quadrierten euklidischen Distanz berechneten Distanzwerte gleichermaßen berücksichtigt werden (vgl. Backhaus et al. 2018: 465f.).[175]

Im Anschluss daran wurden zu den drei Zeitpunkten (TP1, TP2 und TP3) je-

173 Zu *z-scores* normalisierte Werte drücken Datenpunkte anhand ihrer Standardabweichung aus. Es lassen sich folglich Daten vergleichen, die auf verschiedenen Skalenniveaus skaliert sind. Solche *z-scores* lassen sich berechnen, indem ein Datenpunkt vom Mittelwert aller Datenpunkte in einer Datenreihe abgezogen und dann das Ergebnis durch die Standardabweichung der Datenreihe dividiert wird. Dies wird für jeden einzelnen Datenpunkt innerhalb der Datenreihe wiederholt. Eine so normalisierte Datenreihe hat einen Mittelwert von 0 und eine Standardabweichung von 1 (vgl. Field 2018: 36f.).

174 Die quadrierte euklidische Distanz ist ein geometrisches Distanzmaß zwischen zwei Punkten. Zur Berechnung werden die Distanzwerte bzw. die Differenzwerte zwischen zwei Punkten berechnet und quadriert. Die quadrierte euklidische Distanz zählt als Maß der Unähnlichkeit zwischen Objekten bzw. Datenpunkten (vgl. Backhaus et al. 2018: 450).

175 Das Ward-Verfahren ist eines von mehreren hierarchischen *Linkage*-Verfahren zur Bildung von Clustern. Hierarchische Verfahren gehen entweder von allen Subjekten (hier Lernenden) einzeln aus und fassen diese agglomerativ in Gruppen zusammen oder sie beginnen mit allen Subjekten in einer Gruppe, die dann divisiv in kleinere Gruppen aufgeteilt wird. Das Ward-Verfahren zählt zu den agglomerativen Verfahren (vgl. hierzu Backhaus et al. 2018: 456f). Die Autorinnen und Autoren geben an dieser Stelle eine sehr klare Übersicht der verschiedenen Verfahren. Im Gegensatz zu anderen *Linkage*-Verfahren werden nicht diejenigen Subjekte zu „Gruppen zusammengefasst [...], die die geringste Distanz aufweisen, sondern es werden die Objekte (Gruppen) vereinigt, die ein vorgegebenes Heterogenitätsmaß am wenigsten vergrößern. Das Ziel des Ward-Verfahrens besteht folglich darin, jeweils diejenigen Objekte (Gruppen) zu vereinigen, die die Streuung (Varianz) in einer Gruppe möglichst wenig erhöhen. Dadurch werden möglichst homogene Cluster gebildet" (ebd.: 465).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

weils k-means Clusteranalysen mit maximal 500 Iterationen durchgeführt, um die Clusterzugehörigkeit der Teilnehmenden zu den einzelnen Zeitpunkten zu determinieren.[176] Um dem longitudinalen Charakter der Daten und dem dynamisch-komplexen Charakter von Sprachentwicklung gerecht zu werden, bilden in der Analyse die Clusterzentren (*centroids*) des vorherigen Zeitpunkts die initialen Clusterzentren des darauffolgenden Zeitpunkts (Pinel / Csizér 2015: 173). Die Ergebnisse zu jedem Zeitpunkt bauen somit direkt auf den Ergebnissen bzw. der Sprachentwicklung bis zu diesem Zeitpunkt auf. Lernende werden jeweils den Clustern zugeordnet, deren Clusterzentren sie am ähnlichsten sind bzw. dem Clusterzentrum mit der geringsten Distanz, die mithilfe der quadrierten euklidischen Distanz berechnet wird. Die Ergebnisse der Clusteranalysen werden im Anschluss daran in typische Verlaufskurven transformiert. Während die Ergebnisse der einzelnen Cluster Lernende zusammen gruppieren, die sich zu einem bestimmten Zeitpunkt ähnlich verhalten haben, gruppieren die Verlaufskurven Lernende zusammen, die sich im Verlauf des Sprachkurses ähnlich entwickelt haben (vgl. Pinel / Csizér 2015: 173; vgl. für eine schematische Übersicht über den Ablauf der hier verwendeten Kombination aus Clusteranalysen Abbildung 11).

In Abbildung 11, die einen schematischen Ablauf einer longitudinalen Clusteranalyse zeigt, sind einerseits die zu den einzelnen Zeitpunkten gebildeten Cluster zu sehen (schwarze Kästen) und andererseits die Wege, die die einzelnen Teilnehmenden im Zeitverlauf durch diese Cluster verfolgen könnten (farbige Pfeile). Grüne Pfeile zeigen an, dass es zu keinem Wechsel der Cluster kommt. Rote Pfeile verdeutlichen einen Clusterwechsel ‚nach unten', also etwa von Cluster 2 in Cluster 3. Die Pfeile in lila verdeutlichen einen Wechsel ‚nach oben', also etwa einen Wechsel von Cluster 3 in Cluster 2 zwischen den Zeitpunkten 1 und 2. Longitudinale Clusteranalysen erlauben durch diesen Aufbau, spezifische Profile der Lernenden bezüglich ihrer syntaktischen Fertigkeiten herauszuarbeiten und Gruppen zu bilden. Gleichzeitig können Teilnehmende diese Gruppen im Verlauf des Sprachkurses wieder wechseln. Ziel eines solchen Vorgehens war es, herauszuarbeiten, ob es Gruppen von Lernenden gab, die sich in ihrer syntaktischen Progression über die ersten sechs Monate des L2-Erwerbs ähneln und welche verschiedenen Entwicklungskurven sich für einzelne Lernende beschreiben lassen.

Rosmawati (2020) bezeichnet es als *best practice*, syntaktische Komplexität gleichzeitig anhand von verschiedenen Teilkonstrukten wie der Subordination

176 Eine k-Means-Clusteranalyse ist ein iterativer Prozess, d.h., ein Prozess der unter Umständen mehrfach wiederholt bzw. iteriert wird und bei dem mit jeder neuen Iteration die optimalen Clusterzentren bestimmt werden. Die Iterationen werden fortgeführt, bis eine vorher bestimmte Anzahl an Iterationen (500 in diesem Fall) erreicht ist oder die Clusterzentren konvergieren (vgl. Murakami 2014: 143; Backhaus et al. 2018).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

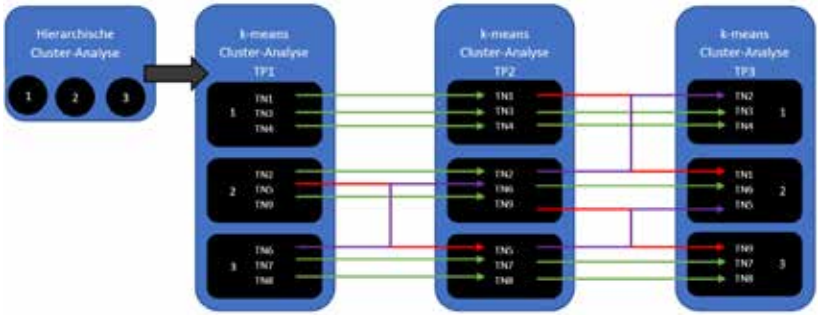


Abbildung 11: Schematischer Ablauf einer longitudinalen Clusteranalyse mit vorgeschalteter hierarchischer Clusteranalyse zur Bestimmung der Clusterzahl

oder der zunehmenden Entwicklung phrasaler Elemente zu untersuchen (vgl. Rosmawati 2020: 114; vgl. Norris / Ortega: 574). Gleichzeitig existieren mittlerweile eine Reihe von Komplexitätsmaßen, die im Bereich der Zweitspracherwerbsforschung Anwendung finden und sich spezifisch auf syntaktische Komplexität beziehen (vgl. Kapitel 2.2). Für die longitudinale Clusteranalyse können aufgrund der Stichprobengröße allerdings nur drei Clustervariablen ausgewählt werden.[177] Vyatkina (2013) und Lu (2011) zeigen, dass folgende Variablen in longitudinalen Studien in hohem Maße distinktiv sind: *mean length of clause* (MLC), *dependent clauses per clause* (DC/C), *complex nominals per clause* (CN/C).[178] In der vorliegenden Studie werden daher diese drei syntaktischen Komplexitätsmaße (vgl. Bulté / Housen 2012 und Kapitel 2.2) adaptiert und ihre Auswahl im Folgenden begründet.[179] Dabei ist zu beachten, dass sich diese Komplexitätsmaße auf die Ebene der minimalen satzwertigen Einheiten (MSE) beziehen. Die in der vorliegenden Studie eingesetzten syntaktischen Komplexitätsmaße sind

- die durchschnittliche Länge minimaler satzwertiger Einheiten (MSE),
- die Subordinationsrate pro MSE,

177 Dolnicar (2002) schlägt für Clusteranalysen in Anlehnung an Formann (1984) eine bevorzugte Stichprobengröße von $5*2^k$ vor, wobei *k* sich auf die Anzahl der verwendeten Clustervariablen bezieht. Das bedeutet für eine Stichprobengröße von 45 Teilnehmenden eine maximale Anzahl von drei Clustervariablen.

178 Neben dem englischsprachigen Forschungskontext werden fast identische Konstrukte in der deutschsprachigen Schreibforschung verwendet (vgl. Petersen 2014: 82–96).

179 Es wäre sicherlich für weitere Forschungsarbeiten mit den zugrundeliegenden Daten interessant, ‚semantische' Variablen zu untersuchen, etwa Variablen, die sich auf den sich entwickelnden Wortschatz der Teilnehmenden beziehen, z. B. die Wortlänge (vgl. Wolfe-Quintero et al. 1998; vgl. Verspoor / Schmid / Xu 2012: 252) oder eine Zählung der verwendeten Lemmata. Diese Auszählungen könnten mit Frequenzlisten des Deutschen abgeglichen werden, um ein exakteres Bild der Wortschatzentwicklung zu erhalten.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

– und die Rate komplexer Nominalphrasen (NPs) pro MSE.

Für diese drei Variablen lassen sich folgende Arbeitsdefinitionen festhalten:

– **Durchschnittliche Länge minimaler satzwertiger Einheiten (MLC; *Mean Length of Clause*)**[180]
Die Anzahl aller Wörter im Gesamttext wird durch die Anzahl minimaler satzwertiger Einheiten (*clauses*) in diesem Text geteilt. Dafür wurden die zuvor im Rahmen der Profilanalyse identifizierten minimalen satzwertigen Einheiten genutzt. Es wurden alle von den Teilnehmenden produzierten Wörter mitgezählt. Dazu zählen etwa Eigennamen und Ortsbezeichnungen ebenso wie Zahlwörter.[181]

– **Subdordinationsrate pro MSE (DC/C; *Dependent Clause Ratio*)**
Die Anzahl aller abhängigen minimalen satzwertigen Einheiten in einem Text wird durch die Anzahl aller minimalen satzwertigen Einheiten in diesem Text geteilt. Dafür wurden die zuvor identifizierten Minimaleinheiten genutzt. Bruchstücke, d.h. Einheiten auf der Stufe 0, wurden aus der Analyse ausgeschlossen. Alle satzwertigen Einheiten, die auf den Stufen 4 (VEND) oder 5 (Integrierter Nebensatz) realisiert wurden, wurden als abhängige MSE gezählt. Obligatorische Kontexte gingen nicht in die Zählung dieser Variable ein.

– **Komplexe Nominalphrasen pro MSE (ComNP/C; *Complex Nominals Ratio*)**
Die Anzahl der komplexen Nominalphrasen in einem Text wird durch die Anzahl an minimalen satzwertigen Einheiten in diesem Text geteilt. Dabei wurde in der vorliegenden Arbeit eine sehr enge Definition komplexer Nominalphrasen gewählt (vgl. für eine Erklärung den folgenden Abschnitt). Im Anschluss an die deutschsprachige Schreibforschung wurden alle Nominalphrasen, die über „minimale Nominalphrasen" (Zifonun et al. 197: 1928f.), also Einwortstrukturen und Phrasen, die nur aus Determinierer und Nomen bestehen, hinausgehen, als komplexe Nominalphrasen beschrieben. Dies umfasste jedoch sehr unterschiedliche komplexe Nominalphrasen. In der vorliegenden Studie wurden daher nur solche komplexen Nominalphrasen ausgezählt, die innerhalb einer „Nominalklammer" (Weinreich 2007: 355) auftreten. Dies beinhaltet die Attribuierung eines Nomens durch Adjektive

180 Bei der Bezeichnung der drei Variablen werden zusätzlich die englischen Bezeichnungen angegeben, die vor allem während der Aufbereitung der Daten verwendet wurden und sich in manchen der Abbildungen finden.

181 Gerade bei Längenmaßen müssen die durchgeführten Segmentationsschritte bei der Aufbereitung der Texte transparent gemacht werden (vgl. Pallotti 2015: 124). Pallotti (2015) schlägt in diesem Zusammenhang eine Operationalisierung von Komplexitätsmaßen vor, die strikt strukturell ausgerichtet und auf spezifische Maße beschränkt ist (vgl. Pallotti 2015: 120). In der vorliegenden Arbeit sind im Anschluss an Pallotti daher strukturelle Merkmale ausschlaggebend für die Berechnung der drei Komplexitätsmaße.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

und Partizipien, nicht jedoch sich an Nomen anschließende Präpositional- oder Genitivphrasen oder Relativsätze.[182] NPs, die nur aus einem Kopf oder nur aus Kopf+Artikel bestehen, wurden aus der Analyse ausgeschlossen, da sie nach der Definition von Lu (2011) nicht komplex sind. Ebenfalls ausgeschlossen wurden NP in der Anrede (z. B. Hallo M. /TN012) oder der Grußformel am Ende (Viele Grüße/TN012).

Die unterschiedlichen syntaktischen Komplexitätsmaße lassen sich in generelle und spezifische Variablen aufteilen. Die Annahme ist, dass es sich bei generellen Maßen um hybride Annäherungen an die Komplexität von Texten handelt, da gerade Längenmaße von Einheiten wie T-Units, Sätzen, satzwertigen Einheiten oder phrasalen Strukturen von verschiedenen Faktoren beeinflusst werden und sich nicht ausschließlich auf die syntaktische Ebene beziehen (vgl. Bulté / Housen 2012: 36). Sie messen letztlich unterschiedliche Konstrukte gleichzeitig. Ebenfalls zu den generellen syntaktischen Komplexitätsmaßen zählen solche, die die Koordination oder die Subordination von Sätzen quantifizieren (vgl. ebd.: 37). Spezifische syntaktische Komplexitätsmaße hingegen beziehen sich konkret auf bestimmte Strukturen und messen beispielsweise die Anzahl von komplexen Nominalphrasen (vgl. für eine Evaluationsstudie verschiedener syntaktischer Variablen Lu 2011; zur Begründung der Auswahl bestimmter Variablen vgl. Vyatkina 2013; vgl. grundlegend Bulté / Housen 2012: 35).

Es bestehen im Deutschen, wie in der Arbeitsdefinition zur ‚Rate komplexer NPs pro MSE' deutlich wurde, zahlreiche unterschiedlich komplexe Möglichkeiten der Nominalphrasenerweiterung. Nominalphrasen bestehen im Deutschen prinzipiell aus einem Nukleus als Kopf der Phrase und aus einem oder mehreren von diesem Kopf abhängigen Elementen. NPs können sowohl durch Rechts- als auch durch Linkserweiterungen zu komplexen Nominalphrasen ausgebaut werden (vgl. Schellhardt / Schroeder 2015: 246–247). Als komplexe Nominalphrase können Strukturen gezählt werden, in denen eine „mehrgliedrige Attribuierung (Ergänzung, Erweiterung) eines Nomens durch Adjektive, Partizipien, Präpositional sowie Genitivattribute und / oder Relativsätze" (Gürsoy 2010:1; vgl. Lu 2010: 483 oder Petersen 2014: 89–96) auftritt. Komplexe Nominalphrasen stellen damit Strukturen dar, in denen es „zu einer Bündelung von Informationen in einer syntaktischen Struktur" (Schellhardt / Schroeder

182 Zur Identifikation von komplexen Nominalphrasen wurde der Stanford Parser für Deutsch (Rafferty / Manning 2008) in der Version 4.2.0 benutzt. Da es sich bei den verwendeten Textbeispielen um (zum Teil normabweichende) Lernersprache handelt, wurden alle identifizierten Nominalphrasen manuell kontrolliert. Wertvolle Hinweise auf weitere Anwendungen und auf Veröffentlichungen im Zusammenhang mit dem Stanford Parser finden sich auf der Projektseite der Stanford Universität unter:
https://nlp.stanford.edu/software/lex-parser.shtml#About [23.02.2023]. Dort finden sich ebenfalls Links zu den neuesten Versionen des Parsers und Hinweise auf Erweiterungen.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

2015: 257) kommt. Diese Art der Informationsverdichtung durch Attribuierung wird als Hinweis auf den Ausbau bzw. auf die Weiterentwicklung von literaten Strukturen in der Lernersprache gedeutet (vgl. ebd.: 246). Eine größere Anzahl an komplexen Nominalphrasen lässt darüber hinaus auf eine höhere „Propositionsdichte" (Petersen 2014: 87) bzw. eine detaillierte Beschreibung der in den Sätzen vorhandenen Propositionen schließen. Dabei wird angenommen, dass sich im Ausbau der Nominalphrasen unterschiedliche Stufen zunehmender struktureller Komplexität unterscheiden lassen.[183]

Die Auswahl der drei Komplexitätsmaße erfolgte für die vorliegende Studie demnach aufgrund folgender Überlegungen:

1. Die ausgewählten Variablen sollten möglichst unterschiedlich sein, d.h. sie sollten unterschiedliche syntaktische Konstrukte messen, um zu garantieren, dass sich unterscheidbare Cluster bilden lassen. Sie durften keine bzw. nicht zu starke Kollinearität aufweisen (vgl. Chan et al. 2015: 306; siehe Tabelle 12).
2. Da die Profilanalysen auf minimalen satzwertigen Einheiten aufbauen (*clauses*), die Variable ‚Erreichte Profilstufe' selbst aber nicht in die Clusteranalyse aufgenommen wurde, da es sich um eine Variable handelt, die in der vorliegenden Studie aufgrund der vergleichsweise wenigen Profilstufen nur sehr wenig Varianz in den Daten versprach, sollten vor allem Variablen ausgewählt werden, die sich auf die *clause*-Ebene beziehen und bei denen größere Varianz in den Daten erwartbar war. Lu (2011) zeigt außerdem, dass Quotienten auf der Grundlage von *clauses* besser zwischen verschiedenen Lernenden diskriminieren können als auf Satzebene oder auf der Ebene von T-Units (vgl. Lu 2011: 56f.).
3. Es sollten generelle Syntaxvariablen und spezielle Syntaxvariablen verwendet werden.

Diese drei Variablen wurden einer Korrelationsanalyse unterzogen, um sie auf Kollinearität zu überprüfen. In der vorliegenden Stichprobe korrelierten diese drei Variablen nur sehr schwach und nicht signifikant miteinander (vgl. Tabelle 12). Alle drei Variablen beziehen sich strukturell auf die *clause*-Ebene und sie repräsentieren die verschiedenen Ebenen: Generelle Syntaxvariable (MLC); Subordination (DC/C) und spezifische Strukturen auf phrasaler Ebene (ComNP/C).

Korrelationen				
		TP1_MLC	TP1_DC/C	TP1_ComNP/C

183 Schellhardt / Schroeder (2015) identifizieren vier Stufen von Nominalphrasen. Die in der vorliegenden Arbeit operationalisierte Definition von Nominalphrasen stimmt mit der zweiten Stufe dieser Klassifikation überein (vgl. Schellhardt / Schroeder 2015: 248–249).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Korrelationen						
TP1_MLC	Korrelation nach Pearson			1	0,213	0,275
	Signifikanz (2-seitig)				0,164	0,070
	N			44	44	44
	Bootstrapc	Verzerrung		0	0,001	0,002
		Standard Fehler		0	0,129	0,131
		BCa 95 % Konfidenz-intervall	Unterer Wert	.	–0,043	0,010
			Oberer Wert	.	0,471	0,522
TP1_DC/C	Korrelation nach Pearson			0,213	1	0,143
	Signifikanz (2-seitig)			0,164		0,356
	N			44	44	44
	Bootstrapc	Verzerrung		0,001	0	–0,010
		Standard Fehler		0,129	0	0,132
		BCa 95 % Konfidenz-intervall	Unterer Wert	–0,043	.	–0,112
			Oberer Wert	0,471	.	0,360
TP1_ ComNP/C	Korrelation nach Pearson			0,275	0,143	1
	Signifikanz (2-seitig)			0,070	0,356	
	N			44	44	44
	Bootstrapc		Verzerrung	0,002	–,010	0
			Standard Fehler	0,131	0,132	0
	BCa 95 % Konfi-denzintervall		Unterer Wert	0,010	–0,112	.
			Oberer Wert	0,522	0,360	.
c. Sofern nicht anders angegeben, beruhen die Bootstrap-Ergebnisse auf 1000 Bootstrap-Stich-proben						

Tabelle 12: Korrelationen von MLC, DC/C und ComNP/C. MLC = Mean Length of Clause → Durchschnittliche Länge minimaler satzwertiger Einheiten (MSE); DC/C = Dependent Clause Ratio → Subordinationsrate pro MSE; ComNP/C = Complex Nominal Phrases Ratio→ Rate komplexer Nominalphrasen pro MSE.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Die Korrelationsanalysen mit 1000 Bootstraps und Bias-corrected and accelerated-Methode (BCa) zeigen folgende Ergebnisse[184]:

Für die Korrelationsanalyse der Variablen ‚TP1_MLC' und ‚TP1_DC/C' zeigt sich keine signifikante Korrelation und nur eine geringe Effektstärke, $r = 0{,}213$ BCa 95 % [−0,043; 0,471], $p = 0{,}164$.[185]

- Die Korrelationsanalyse der Variablen ‚TP1_MLC' und ‚TP1_CompNP' zeigt keine signifikante Korrelation und ebenfalls eine kleine Effektstärke mit $r = 0{,}275$ BCa [0,010; 0,522], $p = 0{,}070$. Hier wird fast das angelegte α-Niveau erreicht. Die Effektstärke bleibt gering, weshalb hier ebenfalls von einer nicht-signifikanten Beziehung zwischen den Variablen ausgegangen werden kann.
- Wird die Korrelation zwischen den Variablen ‚TP1_DC/C' und ‚TP1_CompNP' berechnet, so ergibt die Korrelationsanalyse ebenfalls eine nicht-signifikante Beziehung mit einer kleinen Effektstärke, $r = 0{,}142$ BCa [−0,112; 0,360], $p = 0{,}356$.

Da die drei hier ausgewählten Komplexitätsmaße nur einen Bruchteil der Syntaxentwicklung widerspiegeln, muss eine solche Verwendung spezifischer Komplexitätsmaße genau motiviert und begründet werden (vgl. Bulté / Housen 2012: 36). Hinsichtlich der drei oben genannten Überlegungen lässt sich die Auswahl der drei Variablen statistisch damit legitimieren, dass alle drei Variablen nicht signifikant miteinander korrelieren und sich zusätzlich nur kleine bis maximal mittlere Effektstärken ergeben. Es kann folglich davon ausgegangen werden, dass sie unterschiedliche Konstrukte in den Texten der Lernenden messen.

Zur Überprüfung, inwiefern die experimentelle Manipulation unter Umständen die Ergebnisse verzerren könnte, wurden zwei Analysen durchgeführt. Einerseits zweifache ANOVAs mit Messwiederholung (*2-way repeated measures ANOVA*) anhand der drei Clustervariablen. Für die ANOVAs wurden jeweils die drei Zeitpunkte als Faktoren der Messwiederholung gewählt (*repeated measures factors*). Die Zuteilung zu einer der beiden Instruktionsgruppen (Syntax / Semantik) dient als Zwischensubjektfaktor (*Between-Subject-Factor*). Zusätzlich wurden die Effektstärken berechnet. Für alle Analysen wurden Post-hoc-Tests mit Bonferroni-Korrektur durchgeführt, die die einzelnen Faktoren miteinander vergleichen. Durch einen Chi-Square-Test wurden zuletzt die longitudinalen Cluster daraufhin überprüft, ob die Verteilung der Instruktionsgruppen innerhalb der longitudinalen Cluster statistisch signifikante Unterschiede zeigt.

184 Für eine kurze Erklärung dieser Methode vgl. Fußnote 151.
185 Als Effektstärke wird hier der Korrelationskoeffizient *r* nach Pearson angegeben.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

3.2.6 Zusammenfassung des methodischen Vorgehens bezüglich der Gruppendaten

In diesem Kapitel werden die methodischen Schritte dargestellt, die auf Gruppenebene durchgeführt werden, um einerseits herauszuarbeiten, welche Teilnehmende den beschriebenen Erwerbssequenzen folgen und wie sich die Gruppe als Ganzes verhält. Andererseits kann die syntaktische Entwicklung einzelner Lernender genauer beschrieben werden. Zunächst wurde aufgrund vollständig vorhandener Datensets zu den drei hier untersuchten Zeitpunkten (A1/A2-Test, A2-Test und Goethe-Zertifikat B1; jeweils etwa im Abstand von zwei Monaten) eine Teilstichprobe (n=44) aus allen Teilnehmenden der Gesamtstudie ausgewählt. Diese Teilnehmenden wurden mit der Gruppe der übrigen Teilnehmenden verglichen, um sicherzugehen, dass keine Teilstichprobe ausgewählt wurde, die sich hinsichtlich bestimmter Variablen signifikant von den übrigen Teilnehmenden unterscheidet. Während sich für die Variablen ‚Alter', ‚Geschlecht' und ‚Experimentelle Manipulation' keine Unterschiede ergaben, differierten die gebildete Teilstichprobe und die übrigen Teilnehmenden hinsichtlich der Variable ‚Anwesenheit'. Die Teilnehmenden der Teilstichprobe besuchten deutlich mehr Unterricht. Wie in Kapitel 3.1.2 deutlich wurde, handelte es sich bei den Teilnehmenden der Studie insgesamt um Lernende mit hoher Vorbildung. Die Unterschiede in der Anwesenheit könnten in der Folge darauf hindeuten, dass aus dieser besonderen Stichprobe nun zusätzlich sehr lernaffine Lernende ausgewählt wurden, die unter Umständen spezifische Entwicklungsverläufe aufweisen. Dieser Umstand muss bei der Ergebnisinterpretation berücksichtigt werden, wenn die Ergebnisse in einem größeren L2-Erwerbskontext diskutiert werden sollen.

In den durchgeführten Analysen wurden Verfahren wie die Profilanalyse durch longitudinale Clusteranalysen ergänzt. Diese stellen ein Verfahren dar, dass m. W. bisher noch nicht in der Erforschung des Deutschen als L2 und in Bezug auf syntaktische Variablen verwendet wurde.

3.3 Methodisches Vorgehen: Einzelfallanalysen

Im vorherigen Kapitel wurde gezeigt, mit welchen Methoden Entwicklungen auf Gruppenlevel gezeigt werden können. Um zu verdeutlichen, wie dynamisch einzelne zweitsprachliche Entwicklungen ablaufen können, wird im folgenden Teil das methodische Vorgehen einer Einzelfallanalyse der Entwicklungsverläufe von zwei sehr ähnlichen Lernenden vorgestellt.[186] Einzelfallstudien mit einer oder mehreren Lernenden sind geeignet, um elizitierte Sprachdaten ausgewähl-

186 Ähnliche longitudinale Einzelfallstudien (*Single-Case-Studies*), die nur wenige Lernende über einen längeren Zeitraum in den Blick nehmen, sind etwa Spoelmann / Verspoor (2010), MacQueen (2012), Baba / Nitta (2014) oder Czinglar (2014) (vgl. dazu Kapitel 2.4 und 2.5 im Forschungsstand).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

ter Lernender sehr differenziert und kleinteilig zu analysieren und Lernende miteinander zu kontrastieren. Darüber hinaus ermöglichen es Einzelfallstudien im Gegensatz zu Studien mit einer großen Anzahl von Lernenden, intra- und interindividuelle Variation sichtbar zu machen und damit den Blick für verschiedene Entwicklungsmuster und -prozesse zu öffnen (vgl. Duff 2014: 235).

In diesem Fall ging es um zwei Geschwister, die sich in sehr vielen Hintergrundvariablen ähnelten und zusätzlich innerhalb der Studie über die gesamte Studiendauer dieselbe Kursgruppe besucht haben. Die Analysen beziehen sich auf die Forschungsfragen 1 und 2, wobei Forschungsfrage 2 sich in diesem Analyseteil nicht nur auf die ersten sechs Monate der Studie, sondern den gesamten Studienverlauf bezieht:

Forschungsfragen zur Entwicklung syntaktischer Kompetenz über die Zeit

Forschungsfrage 1: *Welche syntaktische(n) Erwerbsreihenfolge(n) ergibt (ergeben) sich für erwachsene Deutschlernende mit Arabisch als L1?*

Forschungsfrage 2: *Wie lässt sich die syntaktische Entwicklung von ausgewählten Teilnehmenden über die gesamte Studiendauer (18 Monate) charakterisieren?*

Der Fokus lag in diesem Teil der Untersuchung folglich nicht so sehr auf dem letztlich erreichten Niveau, der erreichten syntaktischen Profilstufe oder dem sprachlichen Endprodukt als solchem, sondern auf der dynamischen Entwicklung syntaktischer Strukturen und der Interaktion zwischen verschiedenen Faktoren, die den Entwicklungsprozess beeinflussen (vgl. Chan 2015: 66). In diesem Abschnitt werden zunächst die beiden Geschwister in Bezug auf ihre demografischen Hintergrunddaten genauer vorgestellt und es wird erklärt, warum sich die beiden Geschwister gut für Einzelfallanalysen eignen (vgl. Kapitel 3.3.1). Daran anschließend wird die Textauswahl für die vorliegenden Einzelfallanalysen begründet (vgl. Kapitel 3.3.2). Im Folgenden werden die in diesem Teil der Studie verwendeten Methoden vorgestellt. Bei diesen handelt es sich um Variabilitäts- und Trendanalysen auf der Grundlage von Rohdaten und gleitenden Durchschnitten (vgl. Kapitel 3.3.3), Korrelationsanalysen auf der Grundlage genereller und gleitender Korrelationen (vgl. Kapitel 3.3.4) und *Change Point*-Analysen (CPAs) zur Identifizierung signifikanter Veränderungen in den Zeitreihendaten der beiden Lernenden (vgl. Kapitel 3.3.5). In den in diesem Abschnitt durchgeführten Analysen werden verschiedene longitudinale Sprachdaten (‚Erreichte Profilstufe', ‚Durchschnittliche Länge minimaler satzwertiger Einheiten', ‚Subordinationsrate', ‚Rate komplexer Nominalphrasen pro MSE') der beiden Teilnehmenden verglichen.

3.3.1 Beschreibung

Bei den beiden Geschwistern handelte es sich um die Teilnehmenden TN024 und TN025. Zu Beginn der Studie war TN024 28 Jahre alt und TN025 22 Jahre alt. Sie haben dasselbe Geschlecht. Die Familie der beiden Geschwister stammt aus

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

einer Stadt im Süden Syriens. Beide besuchten in Syrien die Schule und schlossen dort in ihrer Heimatstadt ihre Schulbildung ab. Nach Abschluss der Schule besuchte TN024 für drei Jahre die Universität. TN024 gab an, dass er einen Studiengang im Bereich Wirtschaft mit einem Schwerpunkt in Buchhaltung und Buchhaltungswesen besuchte. Als TN025 sein Studium aufnehmen wollte, begann in Syrien der Bürgerkrieg. Beide erreichten Deutschland Ende 2015, nach einer langen Flucht mit einem Aufenthalt in Jordanien. TN024 hat an der Universität ein wenig Englisch gelernt. TN025 gab keine weiteren Fremdsprachenkenntnisse an. Die zwei Teilnehmenden wuchsen seit ihrer Geburt mit dem syrischen Dialekt auf, der ebenfalls im familiären Umfeld der beiden gesprochen wurde. Hier gaben beide sehr hohe Werte innerhalb der Selbstauskunft des Fragebogens zum Sprachhintergrund und zur Sprachbiografie an. TN024 und TN025 bewerteten ihre Fähigkeiten im Dialekt auf einer 10-Punkt-Skala in den Bereichen Lesen, Hören, Schreiben, Sprechen und grammatische Fertigkeiten jeweils mit 10 Punkten. Das Moderne Standard Arabisch (MSA) lernten beide ab dem Alter von sechs Jahren in der Schule. TN024 bewertete die eigenen Fähigkeiten durchgehend mit 10 Punkten, während TN025 sich bei der Fertigkeit Lesen im MSA mit 7 Punkten, beim Hören mit 8 Punkten und beim Schreiben mit 7 Punkten bewertete. Im Bereich Sprechen und bei den grammatischen Fertigkeiten bewerteten sich TN025 nur mit 5 Punkten.

Beide kamen Ende 2015 gemeinsam in Deutschland an. Sie erlebten die gesamte Flucht nach Deutschland gemeinsam.[187] Vor dem Sprachkurs am MPI CBS besuchten beide keinen organisierten Sprachkurs und lernten nur ungesteuert ein wenig Deutsch in einer Leipziger Sammelunterkunft. Beide TN besuchten in den ersten sechs Monaten Kursgruppe 3, der ein semantischer Schwerpunkt zugrunde lag. In der zweiten Hälfte der Studie besuchten die Teilnehmenden ebenfalls den Kurs mit Schwerpunkt Semantik.

Es wird deutlich, dass sich die beiden Geschwister in vielerlei Hinsicht ähneln und daher in besonderem Maße für einen Vergleich innerhalb einer Einzelfallstudie in Frage kommen.[188] Die Ähnlichkeit der beiden Teilnehmenden ermöglichte so einen sehr kleinteiligen Vergleich. Zu den Ähnlichkeiten, die sich aufgrund der biografischen Daten und der Kurszugehörigkeit ergaben, kamen vergleichbare Hintergrunddaten, die sich im Rahmen der Datenerhebung innerhalb der Studie ergeben. Beide Lernenden nahmen an beinahe allen Tests teil,

187 Es scheint jedoch vor dem Hintergrund der in Kapitel 3.1.2 besprochenen Umstände als unsicher, ob durch die gemeinsame Flucht direkt auf eine gemeinsame Erfahrungswelt im Kontext dieser Flucht geschlossen werden kann.

188 Eine ähnliches methodisches Design findet sich in Czinglar (2014). Oft werden in Einzelfallstudien Zwillinge in ihrer Entwicklung betrachtet um eine größtmögliche Vergleichbarkeit zwischen Lernenden zu gewährleisten (vgl. Chan 2015; vgl. Chan et al. 2015: 301 für eine Übersicht über weitere Zwillingsstudien aus dem englischsprachigen SLA-Kontext).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

wodurch fast vollständige und sehr dichte Datensets zur Verfügung stehen. Darüber hinaus verpassten beide fast keinen Unterricht und haben durchgehend sehr hohe Anwesenheitsquoten, was die tatsächliche inhaltliche Progression im Unterricht hinsichtlich der beobachteten Entwicklungen für die Analysen dieser Arbeiten noch relevanter erscheinen lässt.

Bis zum Ende der ersten Studienphase besuchte TN024 560 von insgesamt 615 möglichen Unterrichtseinheiten, was einer Anwesenheitsquote von 91 % entspricht. TN025 besuchte im selben Zeitraum 592 von 615 Unterrichtseinheiten, was einer Anwesenheitsquote von 96 % entspricht. Beide Anwesenheitsquoten liegen deutlich über dem Durchschnitt für den Rest der Teilnehmenden und sind vor dem Hintergrund der vielfältigen Stresssituationen im Kontext der Ankunft in Deutschland mehr als beachtlich.[189] Alle hier beschriebenen demografischen Hintergrunddaten sind im Folgenden in Tabelle 13 zusammengefasst.

Demografische Hintergrundvariable	**TN024**	**TN025**
Alter zu Beginn der Studie *(AoO)*	28	22
Jahre in Bildungskontexten		
Schulabschluss	Ja	Ja
Universitäre Bildung	Ja	Nein
Universitärer Abschluss	Ja	Nein
Studienrichtung	Wirtschaft	
Ankunft in Deutschland	Ende 2015	Ende 2015
Sprachbiografie		
L1	Arabisch (syrischer Dialekt, Süden)	Arabisch (syrischer Dialekt, Süden)
Erster Kontakt mit L1	Seit der Geburt	Seit der Geburt
Erster Kontakt mit MSA	Im Alter von sechs Jahren (Schulbeginn)	Im Alter von sechs Jahren (Schulbeginn)
Weitere Sprachkenntnisse	Englisch	Keine
Selbsteinschätzung Arabischkenntnisse (max. 10 Punkte)		
Lesen – Dialekt	10	10
Lesen – MSA	10	7
Hören – Dialekt	10	10
Hören – MSA	10	8

189 Die Anwesenheitsquote aller Teilnehmenden für die erste Studienphase bis zum Goethe-Zertifikat B1 liegt bei etwa 58 %. Hier gingen die Abbrecherinnen und Abbrecher ebenfalls in die Berechnung ein. Am Ende der Studie besuchten die 42 Teilnehmenden, die an der zweiten Studienphase teilnahmen, im Schnitt 1175 Unterrichtseinheiten. Dies sind deutlich weniger UE als TN024 und TN025 im Rahmen der Studie besuchten.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Demografische Hintergrundvariable	TN024	TN025
Schreiben – Dialekt	10	10
Schreiben – MSA	10	7
Sprechen – Dialekt	10	10
Sprechen – MSA	10	5
Grammatische Fertigkeiten – Dialekt	10	10
Grammatische Fertigkeiten – MSA	10	5
Anwesenheit innerhalb der Studie		
Anwesenheit absolut (1. Studienphase)	560 / 615	592 / 615
Anwesenheit prozentual (1. Studienphase)	91 %	96 %
Anwesenheit absolut (2. Studienphase)	1292 / 1605	1368 / 1605
Anwesenheit prozentual (2. Studienphase)	80 %	85 %

Tabelle 13: Hintergrundvariablen TN024 und TN025

3.3.2 Auswahl der Texte

Insgesamt stehen für die beiden Teilnehmenden zu 40 verschiedenen Zeitpunkten über die Zeit der gesamten Studie schriftliche Texte zur Verfügung. Nach dem Ausschluss von 16 dieser 40 Texte kamen 24 Texte (bzw. Zeitpunkte) für die Analyse in Frage. Dabei sind für TN024 22 dieser Texte und für TN025 23 dieser Texte vorhanden. Für alle diese Texte wurden im Folgenden als Vorbereitung der nächsten Analysen zunächst die Arbeitsschritte der Profilanalyse durchgeführt (Segmentierung der Texte in minimale satzwertige Einheiten, Einstufung dieser MSE gemäß ihrer syntaktischen Struktur und Einstufung des gesamten Textes auf eine syntaktische Erwerbsstufe; für eine genaue Beschreibung der einzelnen Schritte vgl. Kapitel 3.2.4). Analog zu den Analysen auf Gruppenebene ergaben sich aus den Profilanalysen Informationen darüber, welche Satzteile mit welchen Frequenzen in den untersuchten Texten vorkamen. Zusätzlich wurden aus den Texten dieselben Variablen extrahiert, die für die longitudinalen Clusteranalysen verwendet wurden (‚Durchschnittliche Länge einer MSE', ‚Subordinationsrate pro MSE' und die ‚Rate komplexer Nominalphrasen pro MSE'). Da alle diese Daten aus den 24 ausgewählten Texten ermittelt wurden und diese in einer chronologischen Abfolge stehen, handelt es sich bei den ermittelten Daten um Zeitreihendaten (vgl. zum Begriff ‚Zeitreihe' Kapitel 3.1.3). Eine Übersicht über die ausgewählten Texte findet sich nachfolgend in Tabelle 14:

TN	Text	Bezeichnung	Datum
TN024 / TN025	1	Sich vorstellen	26.05.2016
TN024/ TN025	2	Was machen Sie nach dem Deutschkurs?	28.06.2016
TN024 / TN025	3	Was hast du gemacht?	11.07.2016
TN024 / TN025	4	Weg zum Institut	15.07.2016

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

TN	Text	Bezeichnung	Datum
TN024 / TN025	5	Regeln in Deutschland	05.08.2016
TN024	6	Mein Tag im Zoo	16.08.2016
TN024 / TN025	7	A1/A2-Test (Leipzig)	07.09.2016
TN024 / TN025	8	Ferien	19.09.2016
TN024 / TN025	9	Das perfekte Abendessen	30.09.2016
TN024 / TN025	10	Bildergeschichte	31.10.2016
TN024 / TN025	11	A2-Test (Berlin)	14.11.2016
TN025	12	Mein schönster Glücksmoment	05.12.2016
TN024 / TN025	13	B1-Prüfung	12.12.2016
TN025	14	Ferien_2	04.01.2017
TN024 / TN025	15	Lieblingskünstler	24.01.2017
TN024 / TN025	16	Karte	08.03.2017
TN024 / TN025	17	Fluchtgeschichte	16.03.2017
TN024 / TN025	18	Lieblingsfilm	03.04.2017
TN024 / TN025	19	Journal	05.04.2017
TN024 / TN025	20	Als ich 18 war	07.04.2017
TN024 / TN025	21	Mein Stadtteil	26.04.2017
TN024 / TN025	22	Ehe auf Zeit	16.05.2017
TN024 / TN025	23	Frühförderung	19.07.2017
TN024 / TN025	24	Arbeit	28.07.2017

Tabelle 14: Ausgewählte Texte für TN024 und TN025

Die Gründe für den Ausschluss der Texte waren entweder zu kurze Textsamples (etwa die Texte ‚Personenbeschreibung' (29.7.2016) oder ‚Postkarte' (8.9.2016), ‚Globalsimulation' (10.5.2017)), zu stark gesteuerte Aufgabenstellungen bzw. zu viel Input in der Aufgabenstellung, der die Antwort zu sehr auf bestimmte Phrasen und Strukturen festlegte (etwa ‚Antwort an Lupo' (29.9.2016), ‚Heidekönigin' (1.2.2017) oder ‚Leserbrief' (17.3.2017)) und Texte, die nicht narrativ waren, sondern nur ein bestimmtes Textsortenwissen abfragten, etwa in Bezug auf Aufgaben zur Prüfungsvorbereitung (z. B. die Texte ‚Bewerbungsschreiben' (6.1.2017) oder ‚Esskultur' (14.7.2017)). Außerdem wurden Texte ausgeschlossen, für die die Aufgabenstellung nicht mehr vorlag oder rekonstruierbar war (etwa Text ‚Brief an Firma' (3.8.2017), ‚Fremdsprachen' (13.7.2017)). Bei einigen Texten war die Auswahl unsicher, am Ende wurde die Entscheidung getroffen, diese nicht in den Analysen zu berücksichtigen (etwa ‚Freunde zum Chatten gesucht' (10.6.2016), ‚Historisches Ereignis' (13.2.2017)). Eine vollständige Auflistung aller Aufgabenstellungen findet sich im Anhang.

Die letztlich ausgewählten Texte sollten möglichst gleichmäßig über den Verlauf der Studie verteilt sein (vgl. Abbildung 12 und Abbildung 13). Jedes der Dreiecke symbolisiert einen Zeitpunkt, zu dem ein Text produziert wird. In den beiden

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Abbildungen finden sich alle 24 Zeitpunkte. In Abbildung 12 fallen jedoch einige Lücken zwischen den Texten auf, die sich nicht durch den Ausschluss von Texten erklären lassen. Diese werden in Abbildung 13 erklärt (rote Pfeile). Zunächst wurden während der beiden Ramadanmonate 2016 und 2017 keine Texte der Lernenden erfasst (erster und vierter roter Pfeil in Abbildung 13). Der zweite rote Pfeil erklärt die Lücke an erfassten Texten zum Jahreswechsel 2016/2017, zu dem die Kurse für etwas mehr als zwei Wochen pausierten. Die dritte größere Lücke, die mit einem roten Pfeil verdeutlicht wird, ist im Februar 2017 zu sehen. Kurs 5 und 6 wurden Mitte Februar abgeschlossen, der zweite Teil der Studie startete erst Anfang März, wodurch sich diese Lücke bei den erfassten Texten erklären lässt. Auffällig ist, dass im zweiten Teil der Studie drei der erfassten Texte sehr eng beieinanderliegen. Dies wird durch den roten Kreis in Abbildung 13 verdeutlicht. Alle drei Texte wurden in derselben Woche erhoben. Es handelt sich hierbei um die Texte Nr. 18 („Lieblingsfilm"), 19 („Journal") und 20 („Als ich 18 war"). Dieser Umstand muss bei den Auswertungen im nächsten Teil mitbedacht werden.

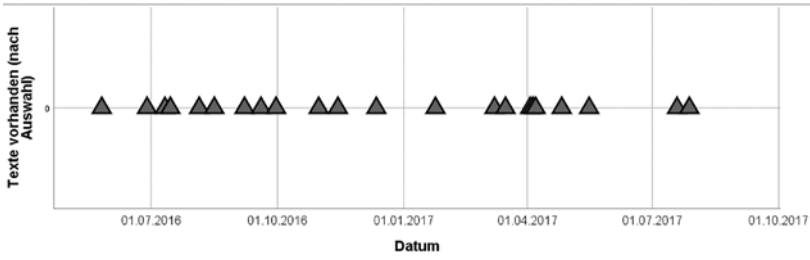


Abbildung 12: Texte zu Zeitpunkten TN024 und TN025

Abbildung 13: Produzierte Texte im Zeitverlauf der Studie. Die roten Pfeile markieren Lücken in der Frequenz der gesammelten schriftlichen Daten.

3.3.3 Methodisches Vorgehen: (Grafische) Trendanalysen

In einem ersten Analyseschritt wurden die Zeitreihendaten (zur Definition von Zeitreihen und Zeitreihendaten vgl. Kapitel 3.1.3 und Backhaus et al. 2018: 127f.) der beiden Geschwister für die untersuchten Variablen mittels verschiedener Diagramme visualisiert. Zunächst wurden die Rohdaten in den Blick genom-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

men. Für die Daten zu den Wortstellungsmustern wurden hier sowohl absolute als auch relative Daten in die Diagramme aufgenommen. ‚Absolut' bezieht sich hier auf die Gesamtzahl der jeweiligen in einem Text produzierten Muster, während ‚relativ' bedeutet, dass der prozentuale Anteil der jeweiligen Wortstellungsmuster an allen produzierten Satztypen dargestellt wird. In einem zweiten Schritt wurden die Daten mithilfe eines gleitenden Durchschnitts (*moving average*) geglättet, um Trends, die den Daten zugrunde liegen, besser sichtbar zu machen. Daten können geglättet werden, indem mehrere Datenpunkte aggregiert und gleichzeitig betrachtet werden. Solche Visualisierungen helfen, eine Richtung bzw. einen Bewegungsvektor in den Daten zu identifizieren (vgl. van Dijk / Verspoor / Lowie 2011: 72). Sie können ebenfalls hilfreich sein, um Höhepunkte und Tiefpunkte mit den Trends in Beziehung zu setzen (vgl. ebd.: 75). Dabei stellen die geglätteten Daten keinen Ersatz für die Rohdaten dar, sondern müssen mit diesen gemeinsam analysiert werden (vgl. ebd.: 72).

Weiterhin spielte in diesem Kontext das Konzept der Variabilität eine entscheidende Rolle. Die in den Daten vorhandene Variabilität sollte grafisch mithilfe von min-max-Diagrammen sichtbar gemacht werden (vgl. für das Vorgehen generell Verspoor et al. 2008). Solche Diagramme sind hilfreich, da sie die Variabilität innerhalb von Daten sichtbar machen und gleichzeitig die Rohdaten weiterhin die Grundlage der Analyse bilden (vgl. van Dijk / Verspoor / Lowie 2011: 76). Zur Erstellung eines min-max-Diagramms wird ein sich bewegendes Fenster an die Daten angelegt, das jeweils einen Datenpunkt weiterbewegt wird und dann innerhalb dieses Fensters den minimalen bzw. den maximalen Wert zurückgibt. Alle in dieser Weise angelegten Fenster überlappen deutlich mit dem vorherigen und dem folgenden Fenster. Ein überlappender Ansatz erlaubt einen detaillierteren Blick auf die beobachteten Muster. Ein Nachteil überlappender Fenster ist, dass große Entwicklungssprünge unter Umständen nicht in ihrer tatsächlichen ‚Größe' wahrgenommen werden (vgl. für eine Diskussion dieses Vorgehens mit überlappenden Fenstern Murakami 2014: 127f.). Aus diesem Grund wurden in der vorliegenden Studie immer sowohl die Rohdaten als auch die durch gleitende Fenster geglättete Daten miteinander verglichen. In den hier durchgeführten Analysen wurde ein Fenster mit dem Intervall ‚5 Zeitpunkte/ Texte' gewählt.[190] Das entscheidende Merkmal eines min-max-Diagramms ist, dass durch die minimale und maximale Grenze die Bandbreite der beobachteten Werte sichtbar gemacht wird. Diese nun sichtbare Bandbreite kann im nächsten Schritt auf größere Schwankungen überprüft werden. Unterschiedliche Phasen können dann in Beziehung zu möglichen Entwicklungssprüngen gesetzt werden, die in dieser Arbeit als *Change Points* (vgl. Kapitel 3.3.5) operationalisiert werden. Umso größer die Bandbreite zwischen den minimalen und

190 Hier sind durchaus unterschiedliche Operationalisierungen der *moving windows* denkbar. Neben unterschiedlichen Zeiteinheiten lassen sich etwa obligatorische Kontexte für *moving windows* operationalisieren, die dann zeitpunktübergreifend berechnet werden (vgl. für dieses Vorgehen Murakami 2014: 125–131).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

maximalen Graphen, desto größer ist die beobachtete Variabilität bzw. Variation in diesem Zeitraum.

Um daran anschließend die Beziehungen zwischen Variablen zu beschreiben, wurde auf deskriptive Operationalisierungen aus dem Kontext der CDST zurückgegriffen (vgl. Kapitel 2.5.1; vgl. Larsen-Freeman 2006: 592–593). Beziehungen können in ihren grundlegenden Verhältnissen zueinander

1. kompetitiv sein, was bedeutet, dass die Werte von Variable a steigen, während die Werte von Variable b kleiner werden oder anders herum.
2. unterstützend sein. Dies ist der Fall, wenn die Werte beider beobachteten Variablen gemeinsam ‚wachsen'.
3. konditionell sein. Als konditionell werden Variablen bezeichnet, wenn das Verhalten einer oder mehrerer Variablen vom Verhalten einer oder mehrerer anderen Variablen abhängt (z. B. ein konditioneller Zusammenhang zwischen Wortschatz und grammatischen Fertigkeiten).

Die Variablen werden grundsätzlich als miteinander verbunden gesehen (*connected growers*) (vgl. dazu Kapitel 2.5.1).

Es ist darüber hinaus im Blick zu behalten, dass es sich hier lediglich um eine Visualisierung der Daten handelt und nicht auf kausale Zusammenhänge oder direkt auf Gründe für die beobachteten Muster geschlossen werden kann. Gleichzeitig bilden Trendanalysen letztlich immer nur individuelle Präferenzen der Lernenden bei der Konstruktion ihrer Texte ab und lassen daher grundsätzlich nur indirekt Schlüsse auf die unterliegende zweitsprachliche Entwicklung zu.

3.3.4 Methodisches Vorgehen: Generelle und gleitende Korrelationen

Zur Überprüfung der in den grafischen Analysen gefundenen Trends wurden KorrelatIonsanalysen durchgeführt, die die Beziehung zwischen den Variablen genauer in den Blick nehmen. Da jedoch von dynamischen Bedingungen zwischen den untersuchten Variablen ausgegangen wurde, wurden die generellen (linearen) Korrelationsanalysen durch gleitende Korrelationsanalysen (*moving correlations*) ergänzt, die jeweils einen bestimmten Ausschnitt der Zeitreihen fokussieren und für diesen einen Korrelationskoeffizienten *r* zurückgeben.[191] Hierbei wurde jeweils ein Korrelationsintervall von fünf Messzeitpunkten angelegt (vgl. für dieses Vorgehen z. B. Spoelmann / Verspoor 2010). Bei dieser Intervallsetzung handelte es sich letztlich um eine arbiträre Setzung. Bulté / Housen (2020) machen darauf aufmerksam, dass ein Intervall von fünf Datenpunkten

191 Einführend zur Technik der gleitenden Korrelationsanalyse ist der Beitrag von Verspoor / van Dijk (2011) zu empfehlen. An den Beitrag ist außerdem eine ausführliche *How to Section* angeschlossen, in der die einzelnen Analyseschritte detailliert erklärt werden.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

unter Umständen zu klein sein könnte, um verlässliche Korrelationskoeffizienten zu berechnen und empfehlen weitere Intervalle zu wählen (vgl. Bulté / Housen 2020: 214). In der Studie von Spoelman / Verspoor (2010), auf die die Kritik sich bezieht, werden deutlich mehr Datenpunkte (N=54) ausgewertet als in der vorliegenden Studie, was letztlich einer der Gründe war, ein kleineres Intervall zu wählen. Bulté / Housen (2020) machen deutlich, dass es bis heute keine allgemein akzeptierte Norm oder Empfehlung bei der Wahl von Intervallgrößen gibt (vgl. Bulté / Housen 2020: 214). Während der Korrelationskoeffizient im Fall von gleitenden Korrelationen gleichzeitig die Effektstärke zurückgibt, wird in den Diagrammen nicht zusätzlich angegeben, ob die jeweilige Korrelation signifikant war bzw. was der *p*-Wert für die gefundene Korrelation war. Hier sollte nicht eine einzelne (nicht) signifikante Beziehung geprüft werden, sondern der Fokus darauf liegen, aufzuzeigen, dass sich die Beziehungen zwischen Variablen im Laufe der Zeit sehr dynamisch ändern und Korrelationen in verschiedenen Studienphasen stärker und schwächer ausgeprägt sein können.

3.3.5 Methodisches Vorgehen: *Change Point*-Analysen

Zur Analyse der in Zeitreihen dargestellten Daten von TN024 und TN025 wurden *Change Point*-Analysen (CPAs) durchgeführt. Diese Art von Analysen wird verwendet, um in Zeitreihen plötzliche und signifikante Veränderungen der Mittelwerte oder der Standardabweichungen (bzw. der Varianz) aufzuspüren (vgl. MacIntyre et al. 2017) und zu entscheiden, ob sich das Verhalten der Zeitreihe vor und nach der Veränderung (dem *Change Point* (CP)) signifikant voneinander unterscheidet (vgl. Taylor 2000|2018). Bei der Erstellung der CPAs in der vorliegenden Arbeit wurde der *Change Point Analyzer* von Taylor Enterprises in der Version 2.3 verwendet.

Das in der vorliegenden Studie verwendete Verfahren zur Durchführung einer CPA operationalisiert einen CUSUM-Algorithmus, der auf einer Kombination aus CUSUM-Diagrammen und Bootstrap-Techniken bereuht. CUSUM steht für *cumulative sum chart* (Kumulativsummen-Diagramm). Bei einem solchen Diagramm werden die „kumulative Summen der Differenzen zwischen den Werten und dem Durchschnitt [addiert]. Diese Differenzen summieren sich zu null, so dass die kumulative Summe immer bei null endet" (Taylor 2000|2018: 5; [Übersetzung MS]). Steigende Werte bzw. eine ansteigende Linie in einem CUSUM-Diagramm deuten daher daraufhin, dass die Werte der ursprünglich untersuchten Variable sich oberhalb des Mittelwertes befinden, während eine abfallende Linie signalisiert, dass die ursprünglichen Werte unterhalb des Mittelwerts liegen. Eine Änderung dieser Tendenzen verdeutlicht daher eine Änderung in der untersuchten Zeitreihe in Bezug auf den Mittelwert, einen *Change Point* (vgl. Abbildung 14).

Mit den aus den CUSUM-Diagrammen berechneten Differenzwerten wird dann eine Bootstrapping-Analyse durchgeführt, wobei durch das Bootstrapping zu-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

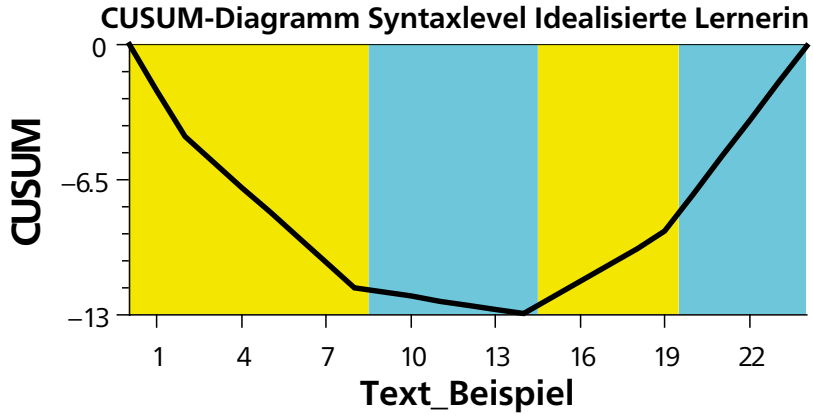


Abbildung 14: CUSUM-Diagramm einer CPA für eine idealisierte Lernerin

fällige Neuanordnungen der Daten hergestellt werden, die die Anordnung der Daten für den Fall nachahmen, dass keine Veränderung in der Zeitreihe stattgefunden hätte. Damit diese Analyse fehlerfrei funktionieren kann, müssen die Daten die Voraussetzung (*assumption*) einer unabhängigen Fehlerstruktur (*independent error structure*) erfüllen. Die einzelnen Datenpunkte in der Zeitreihe dürfen sich folglich nicht überschneiden oder negativ oder positiv miteinander korreliert sein. Falls in der Analyse die Differenzen der Bootstrap-CUSUM-Diagramme dazu tendieren, näher an null zu liegen als die der ursprünglichen Zeitreihe im CUSUM-Diagramm, kann davon ausgegangen werden, dass es an den betreffenden Stellen zu einer Veränderung, einem *Change Point*, gekommen ist (vgl. Taylor 2000|2018: 7). Bei den CPAs in der vorliegenden Arbeit werden immer 1000 Bootstrapping-Analysen durchgeführt. Für diese Bootstrap-Analysen wird dann ein Konfidenzniveau angeben, welches verdeutlicht, wie viele der Iterationen des Bootstraps näher an null lagen als die Daten des ursprünglichen CUSUM-Diagramms. Werden etwa 1000 Bootstrapping-Iterationen durchgeführt, von denen 950 näher an null liegen als die Ausgangsdaten, ergibt sich ein Konfidenzniveau von 95 %.[192] Zuletzt werden zusätzlich CPAs der Variation innerhalb der Daten herangezogen, um einen Blick darauf zu gewinnen, wann die Variation in den Daten sich ändert. Dies ist vor allem aus CDST-Perspektive relevant, da Variation bzw. Variabilität als einer der Auslöser für Veränderung gilt (vgl. Bot / Lowie / Verspoor 2007; Verspoor et al. 2008; Spoelman / Verspoor

192 Um die statistische Signifikanz der auf diese Weise gefundenen *Change Points* genauer zu überprüfen, könnten beispielsweise Monte Carlo-Analysen durchgeführt werden (vgl. zu diesem Punkt außerdem Kapitel 5).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

2010).[193] Steenbeek / Jansen / van Geert (2012) gehen davon aus, dass CPAs dazu verwendet werden können, stabile Zustände in einer Zeitreihe zu identifizieren (vgl. ebd.: 68).

Der Output einer CPA in Bezug auf die erreichte Profilstufe einer Profilanalyse nach Grießhaber wird in Abbildung 15 für eine idealisierte Lernerin dargestellt.

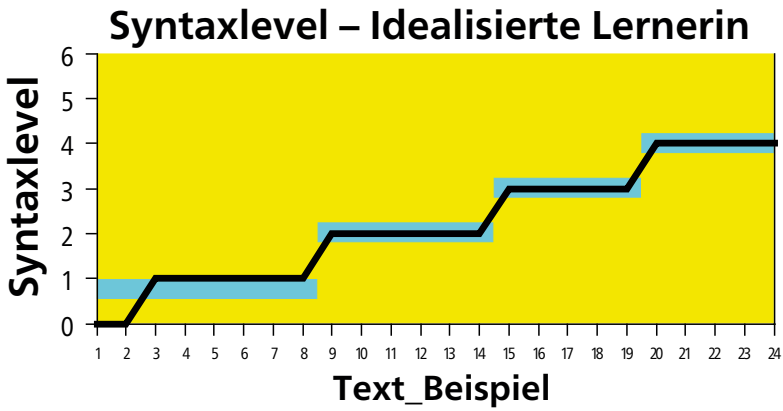


Abbildung 15: Output einer CPA für eine idealisierte Lernerin

In Abbildung 15 sind auf der x-Achse die Nummern von 24 Textsamples angeben. Die y-Achse zeigt die in diesen Texten in einer (fiktiven) Profilanalyse jeweils erreichten syntaktischen Profilstufen an. Die Abbildung verdeutlicht, dass die idealisierte Lernerin nacheinander, d.h. in sequenzieller Weise, die ersten fünf syntaktischen Profilstufen durchläuft. Für jede Profilstufe ergeben sich signifikante CPs, die durch die Verschiebung der hellblauen Balken gekennzeichnet werden. Im vorliegenden Beispiel werden drei CPs zu den Zeitpunkten bzw. Texten 9, 15 und 20 identifiziert. Zusätzlich zum grafischen Output werden die Ergebnisse und die Zusammenfassung der Analyse in einer Tabelle festgehalten (siehe Abbildung 16 unten).

Falls in einer Zeitreihe ein CP identifiziert wird, wird zusätzlich ein Konfidenzniveau angegeben, das einen Hinweis darauf gibt, wie wahrscheinlich es ist, dass tatsächlich ein CP vorliegt (vgl. Taylor 2000|2018: 3). Im idealisierten Beispiel wurde der erste CP mit einem Konfidenzniveau von 100 % identifiziert, für den zweiten wird ein Wert von 98 % und für den dritten ein Wert von 95 % angegeben. Außerdem wird für jeden gefundenen CP ein (zeitliches) Konfidenzinter-

193 Die Analysen der Variation funktionieren analog zu den Analysen der ursprünglichen Zeitreihe. Der Output sieht ebenfalls identisch aus. Deshalb wird an dieser Stelle darauf verzichtet, diese Analyseschritte mit Abbildungen zu verdeutlichen.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Table of Significant Changes for Syntaxlevel

Confidence Level for Candidate Changes = 50%, Confidence Level for Inclusion in Table = 90%, Confidence Interval = 95%, Bootstraps = 1000, Without Replacement, MSE Estimates

Text_Beispiel	Confidence Interval	Conf. Level	From	To	Level
9	(9, 9)	100%	0.75	2	2
15	(15, 15)	98%	2	3	1
20	(20, 20)	95%	3	4	2

Abbildung 16: Tabellarischer Output einer CPA

vall angegeben, welches angibt, wie gut der Zeitpunkt der Änderung bestimmt werden konnte (vgl. Taylor 2000|2018: 4). Dieses liegt im idealisierten Beispiel für den ersten CP nur auf dem 9. Text bzw. Zeitpunkt in der Zeitreihe. Für die anderen beiden identifizierten CPs ergeben sich ebenfalls sehr enge Konfidenzintervalle, was auf die ‚Herstellung' der Daten zurückzuführen ist. Für authentische Lernerdaten ist aufgrund der intraindividuellen Variabilität kaum ein so eindeutiges Ergebnis zu erwarten.

3.3.6 Zusammenfassung des methodischen Vorgehens in Bezug auf die Einzelfallanalysen

In diesem Kapitel werden die methodischen Grundlagen einer longitudinalen Einzelfallanalyse mit zwei eng verwandten Lernenden vorgestellt, die die Analysen auf Gruppenebene ergänzen und vertiefen sollen. Dabei wird die Entwicklung der beiden Geschwister über die gesamte Studiendauer (insgesamt etwa 18 Monate) mithilfe der bereits in den Analysen auf Gruppenebene verwendeten Variablen untersucht. Die beiden Geschwister wurden aufgrund der hohen Vergleichbarkeit, die zwischen den Daten der beiden gegeben ist, ausgewählt, um noch deutlicher Unterschiede und Gemeinsamkeiten in der syntaktischen Entwicklung von sehr ähnlichen Lernenden sichtbar zu machen. Für diese Analysen wurden insgesamt 24 Zeitpunkte ausgewählt, die über die gesamte Studiendauer verteilt sind und so ermöglichen, Entwicklungsprozesse möglichst detailliert und ‚lückenlos' darzustellen.

In einem ersten grundlegenden Analyseschritt wurden hier ebenfalls Profilanalysen für alle Texte der beiden Lernenden durchgeführt, die analog zu den Analysen auf Gruppenebene die Anzahl und die Profilstufen der minimalen satzwertigen Einheiten sowie die Profilstufen der Gesamttexte und zusätzlich die verwendeten Wortstellungsmuster liefern. Analog zu den Gruppenanalysen wurden die Variablen ‚Durchschnittliche Länge einer MSE', ‚Subordinationsrate' und ‚Rate komplexer Nominalphrasen pro MSE' berechnet. Zur Illustration der Entwicklungsprozesse wurden zunächst Trendanalysen durchgeführt und dann durch Korrelationsanalysen und *Change Point*-Analysen (CPAs) vertieft.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Während die Trend- und Korrelationsanalysen in zahlreichen L2-Forschungskontexten verwendet werden, handelt es sich bei CPAs um eine Technik, die in der L2-Erwerbsforschung noch relativ selten und dann hauptsächlich im Kontext der Motivationsforschung angewendet wird. Baba / Nitta (2014) führen verschiedene CPAs durch, um zu untersuchen, wie sich die Selbsteinschätzung während des Erwerbs einer L2 verändert (vgl. Kapitel 2.5.2 zu CDST-basierten empirischen Studien). Hierbei nutzen sie Daten zur Motivation. Zur Analyse syntaktischer L2-Daten des Deutschen werden diese Analysen m. W. bisher nicht genutzt. Es ist daher zusätzlich abzuschätzen, ob diese Analysen für die vorliegenden Daten sinnvolle und hilfreiche Erkenntnisse liefern können.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

4 Ergebnisse

In den nun folgenden Kapiteln werden die Ergebnisse der beiden Studienteile zusammengefasst. In Kapitel 4.1 werden die Ergebnisse der Analysen auf Gruppenebene vorgestellt, wobei zunächst deskriptive Daten zu den Texten der drei Zeitpunkte den Fokus bilden (vgl. Kapitel 4.1.1). Als nächstes werden die Ergebnisse der Profilanalysen (vgl. Kapitel 4.1.2) und der Emergenzanalysen (vgl. Kapitel 4.1.2) präsentiert. Zuletzt werden die Ergebnisse der longitudinalen Clusteranalyse herausgearbeitet (vgl. Kapitel 4.1.3).

An die Ergebnisse der Analyse auf Gruppenebene schließt sich in Kapitel 4.2 die Ergebnisdokumentation der Einzelfallanalysen an. Hier werden ebenfalls zunächst deskriptive Daten gezeigt (vgl. Kapitel 4.2.1). Daran anschließend werden die visuellen Analysen zu Trends und zur Variabilität innerhalb der Daten behandelt (vgl. Kapitel 4.2.2–4.2.4). Die Visualisierungen werden durch die Ergebnisse der Korrelationsanalysen (vgl. Kapitel 4.2.6 und 4.2.7) weiter kontextualisiert. In einem letzten Schritt werden die Ergebnisse der *Change Point*-Analysen (CPAs) vorgestellt und in Verbindung zu den bis zu diesem Punkt durchgeführten Analysen gesetzt (vgl. Kapitel 4.2.8).

Eine Einordung und die Diskussion der Ergebnisse im Kontext bisheriger Forschung erfolgen im nächsten Kapitel (vgl. Kapitel 5).

4.1 Ergebnisse: Gruppendaten

4.1.1 Ergebnisse: Deskriptive Statistiken zu TP1, TP2 und TP3

In die Analyse auf Gruppenebene wurden schriftliche Lernertexte zu drei Zeitpunkten einbezogen (im Folgenden TP1, TP2 und TP3; vgl. zum methodischen Vorgehen in Bezug auf die Textauswahl Kapitel 3.2.3). An dieser Stelle werden für die drei Zeitpunkte deskriptive Statistiken hinsichtlich der Anzahl der in den Texten produzierten Wörter und satzwertigen Einheiten sowie der durchschnittlichen Länge dieser satzwertigen Einheiten (MLC), der Subordinationsrate (DC/C) und der Rate komplexer Nominalphrasen pro minimaler satzwertiger Einheit (ComNP/C) präsentiert, um die Ergebnisse der longitudinalen Clusteranalysen besser einschätzen zu können, in denen anhand der genannten Komplexitätsmaße Cluster zu den drei Zeitpunkten und longitudinale Cluster über den Zeitverlauf gebildet werden.[194] Die deskriptiven Statistiken umfassen jeweils Mittelwerte, Standardfehler der Mittelwerte, Median, Modus, Standardabweichung, Varianz, Minimum und Maximum sowie die Summe der Variablen. Im nächsten

194 Die aus den Lernertexten extrahierten Rohdaten (MSE, obligatorische Kontexte, erreichte Profilstufen etc.) zu den in diesem Kapitel durchgeführten Analysen finden sich im Anhang der Arbeit.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Kapitel (vgl. Kapitel 4.1.2) werden wiederum deskriptive Statistiken berichtet. Diese beziehen sich auf die in den Profilanalysen extrahierten Daten, d.h. auf die produzierten Wortstellungsmuster und die insgesamt erreichte Profilstufe.

Deskriptive Statistik zu Zeitpunkt 1 (TP1):

In Bezug auf die Daten zu den produzierten Wörtern und minimalen satzwertigen Einheiten sowie den Variablen ‚MLC', ‚DC/C' und ‚ComNP/C' zu Zeitpunkt 1 ergibt sich folgendes Bild (vgl. Tabelle 15). Die Werte sind, soweit möglich, jeweils auf die dritte Nachkommastelle gerundet:

Deskriptive Statistik TP1						
		TP1_Wörter	TP1_Satzwertige Einheiten	TP1_MLC	TP1_DC/C	TP1_ComNP/C
N	Gültig	44	44	44	44	44
	Fehlend	0	0	0	0	0
Mittelwert		52,250	10,090	5,158	0,062	0,259
Standardfehler des Mittelwerts		2,303	0,376	0,129	0,013	0,026
Median		53,500	10,000	5,1389	0,000	0,279
Modus		57	10	3,333[a]	0,000	0,000
Std.-Abweichung		15,276	2,495	0,858	0,088	0,168
Varianz		233,355	6,224	0,735	0,008	0,028
Minimum		10	3	3,33	0,000	0,000
Maximum		94	18	7,090	0,300	0,580
Summe		2299	444	226,930	2,740	11,38
Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.						

Tabelle 15: Deskriptive Daten zu den produzierten Texten zu Zeitpunkt 1 (TP1). Gezeigt werden die Daten für die produzierten Wörter, die minimalen satzwertigen Einheiten (MSE), die durchschnittliche Länge der MSE (MLC) in Wörtern, die Subordinationsrate (DC/C) und die komplexen Nominalphrasen pro MSE (ComNP/C).

Die 44 Teilnehmenden produzierten in den 44 Texten zu Zeitpunkt 1 insgesamt 444 satzwertige Einheiten mit insgesamt 2299 Wörtern, was bedeutet, dass jeder Text im Schnitt aus 10,09 satzwertigen Einheiten mit durchschnittlich 52,52 Wörtern und jede satzwertige Einheit durchschnittlich aus etwa 5,2 Wörtern bzw. Token bestand. Schon hier wird deutlich, dass die Teilnehmenden insgesamt sehr kurze Sätze mit wenigen Elementen produzierten. Dies wird durch die relativ geringe Standardabweichung (0,858) der Werte von ‚TP1_MLC' bestärkt. Typische Beispiele mit einer Länge von fünf bzw. sechs Wörtern sind als Belege [1]–[4] aufgeführt[195]:

195 Alle lernersprachlichen Belege werden in der orthografisch nicht korrigierten Form wiedergegeben, in denen sie von den Lernenden produziert wurden. Zusätzlich wird jeweils angegeben, aus welchem Text der jeweilige Beleg stammt.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

[1] *Und wir gehen zu Restaurant.* (TN012; TP1; fünf Wörter bzw. Token)

[2] [...], *dass du nach Leipzig kommst?* (TN038; TP1; fünf Wörter bzw. Token)

[3] *Ich würde dich in Leipzig sehen,* [...]. (TN001; TP1; sechs Wörter bzw. Token)

[4] *Hier können wir im see schwimmen.* (TN041; TP1; sechs Wörter bzw. Token)

Ohne an dieser Stelle in ausreichender Form auf die hier gezeigten Belegstellen eingehen zu können, verdeutlichen sie doch, dass die Länge von Äußerungen kein hinreichender Beleg für zunehmende Komplexität innerhalb von Lernersprache sein kann. Während Beleg [1] auf der Ebene der Satzkonstruktion weniger komplex als Beleg [3] ist, ist dies für den Vergleich der Belege [2], [3] und [4] nicht problemlos festzustellen. Belege [3] und [4] bestehen beide aus sechs Wörtern, werden aber unterschiedlichen Erwerbsstufen zugeordnet. Beleg [4] ist zwar mit fünf Wörtern ein kürzerer Beleg, wird aber der höchsten Erwerbsstufe innerhalb dieser vier Belege zugeordnet. Für die lernersprachlichen Belege in diesem Kapitel werden Belegstellen gewählt, die jeweils nah an den berechneten Mittelwerten der Längenmaße liegen. Dies ist hier und im Folgenden zu beachten, da die Lernenden natürlich MSE produzierten, die zum Teil deutlich kürzer oder länger und gleichzeitig trotzdem syntaktisch gesehen weniger oder stärker komplex sind (vgl. Belege [5] und [6]):

[5] *Es gibt viele Leute,* ***die freundlich sind****.* (TN013; TP1; drei Wörter bzw. Token)

[6] *Wir sehen Viele schauschspeilen, viele Aktivitäten, Viele möbele und Viele Trinken.* (TN044; TP1; elf Wörter bzw. Token)

In diesem extremen Beispiel produzierte TN013 in [5] einen Relativsatz, der aufgrund der Verbendstellung auf Profilstufe 4 eingeordnet wird. Dieser Relativsatz hat gleichzeitig ‚nur' drei Wörter und liegt damit zu diesem Zeitpunkt deutlich unter dem Durchschnitt der produzierten MSE. TN044 produzierte in [6] eine sehr lange MSE mit insgesamt elf Wörtern, die auf der Stufe 1 ‚Finitum' eingeordnet wird. Dabei werden gerade durch solche Beispiele die großen interindividuellen Unterschiede zwischen den Teilnehmenden deutlich. Während ein/-e Teilnehmer/-in einen Text geschrieben hat, der insgesamt nur aus zehn Wörtern bestand, lag das Maximum produzierter Wörter bei 94. Für die produzierten satzwertigen Einheiten lag das Minimum bei 3 und das Maximum bei 18. In den Texten aller Lernenden sind 24 der 444 satzwertigen Einheiten Nebensätze mit Verbendstellung, was die sehr geringen Werte der Subordinationsrate (‚TP1_DC/C') erklärt (für eine Übersicht über die zu diesem Zeitpunkt produzierten Wortstellungsmuster vgl. das folgende Kapitel 4.1.2). Diese liegt zu Zeitpunkt 1 bei 0,062. Eine von 16 satzwertigen Einheiten beinhaltet im Schnitt ein Verb in Verbendstellung. Die Rate komplexer Nominalphrasen pro MSE liegt zu diesem Zeitpunkt bei 0,259, was bedeutet, dass durchschnittlich nur in etwa jeder vierten MSE eine komplexe Nominalphrase produziert wird.

Deskriptive Statistik zu Zeitpunkt 2 (TP2):

Zum Zeitpunkt 2 wurde ein A2-Sprachtest durchgeführt (für eine Beschreibung der zugrundeliegenden Schreibaufgabe aus diesem Test vgl. Kapitel 3.2.3). In

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Tabelle 16 finden sich die deskriptiven Statistiken zu den aus den Texten extrahierten Variablen. Wie in der Tabelle zu Zeitpunkt 1 sind die Werte hier ebenfalls auf die dritte Nachkommastelle gerundet.

Deskriptive Statistik TP2						
		TP2 Wörter	TP2_Satzwertige Einheiten	TP2_MLC	TP2_DC/C	TP2_ComNP/C
N	Gültig	44	44	44	44	44
	Fehlend	0	0	0	0	0
Mittelwert		75,84	13,02	5,880	0,105	0,343
Standardfehler des Mittelwerts		2,356	0,442	0,100	0,018	0,018
Median		72,00	12,50	5,845	0,081	0,310
Modus		62	11	5,64	0,00	0,30[a]
Std.-Abweichung		15,625	2,929	0,664	0,120	0,117
Varianz		244,137	8,581	0,441	0,014	0,014
Minimum		46	8	3,69	0,00	0,13
Maximum		109	20	7,90	0,45	0,55
Summe		3337	573	258,71	4,62	15,08
a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.						

Tabelle 16: Deskriptive Daten zu den produzierten Texten zu Zeitpunkt 2 (TP2)

Zu diesem Zeitpunkt wurden von den 44 Teilnehmenden 3337 Wörter und 573 satzwertige Einheiten produziert. Jeder Text bestand nun im Schnitt aus 13,02 satzwertigen Einheiten mit einer durchschnittlichen Länge von ~5,9 Wörtern. Typische Beispiele von zu diesem Zeitpunkt produzierter MSE mit einer Länge von sechs bis sieben Wörtern finden sich in den Belegen [7]–[10]:

[7] […] *und ich habe den Alexanderplatz besucht.* (TN001; TP2; sechs Wörter bzw. Token)

[8] *Dann haben wir Mittagessen zusammen gegassen.* (TN038; TP2; sechs Wörter bzw. Token)

[9] *Zuerst sind wir mit dem Buss gefahren.* (TN039; TP2; sieben Wörter bzw. Token)

[10] *Meine Familie und ich waren sehr glücklich*, […] TN001; TP2; sieben Wörter bzw. Token)

Auffällig ist, dass für die Variablen ‚Wörter' und ‚Satzwertige Einheiten' die Standardabweichung zu Zeitpunkt 2 im Vergleich zu Zeitpunkt 1 sehr ähnlich bleibt. Für die Variable ‚TP2_MLC' wird sie etwas kleiner. Insgesamt scheint die Streuung der Daten ähnlich zu bleiben wie zu Zeitpunkt 1. Das Minimum produzierter satzwertiger Einheiten liegt bei acht Einheiten und das Maximum bei 20. Ebenso zeigen die Werte der übrigen Variablen deutliche Steigerungen. Die Subordinationsrate liegt bei 0,105. Jede zehnte satzwertige Einheit ist zu diesem

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Zeitpunkt eine Einheit mit Verb in Endstellung. Das Maximum liegt bei 0,45. Manche Teilnehmende produzierten durchschnittlich in jedem zweiten Satz subordinierende Nebensätze. Die Rate komplexer Nominalphrasen steigt auf 0,343. Das Maximum liegt hier bei 0,55.

Deskriptive Daten zu Zeitpunkt 3 (TP3):

Zeitpunkt 3 war der Zeitpunkt des Goethe-Zertifikats B1 (für eine Beschreibung der zugrundeliegenden Schreibaufgabe aus diesem Test vgl. Kapitel 3.2.3). Tabelle 17 gibt einen Überblick über die deskriptiven Statistiken zu den aus den Texten extrahierten Variablen. Wie zuvor sind die Werte hier auf die dritte Nachkommastelle gerundet.

Deskriptive Statistik TP3						
		TP3 Wörter	TP3_Satzwertige Einheiten	TP3_MLC	TP3_DC/C	TP3_ComNP/C
N	Gültig	44	44	44	44	44
	Fehlend	0	0	0	0	0
Mittelwert		102,98	19,23	5,385	0,177	0,322
Standardfehler des Mittelwerts		3,029	0,598	0,071	0,016	0,022
Median		100,50	19,00	5,335	0,182	0,294
Modus		89	17	4,59[a]	0,00	0,29[a]
Std.-Abweichung		20,093	3,964	0,471	0,109	0,149
Varianz		403,744	15,715	0,222	0,012	0,022
Minimum		57	11	4,59	0,00	0,06
Maximum		150	30	6,50	0,47	0,69
Summe		4531	846	236,96	7,80	14,16
a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.						

Tabelle 17: Deskriptive Daten zu den produzierten Texten zu Zeitpunkt 3 (TP3)

Im Vergleich zu Zeitpunkt 2 werden zu Zeitpunkt 3 mehr Wörter (insgesamt 4531) und satzwertige Einheiten (insgesamt 846) von den Teilnehmenden produziert. Während sich die Durchschnittswerte für Wörter und satzwertige Einheiten pro Text weiter erhöhen (die Mittelwerte liegen jetzt bei 102,98 Wörtern pro Text und 19,23 satzwertigen Einheiten pro Text), werden die satzwertigen Einheiten gleichzeitig etwas kürzer. Lag die durchschnittliche Länge zu TP2 noch bei etwa 5,9 Wörtern pro MSE, liegt der Wert zu Zeitpunkt 3 nur noch bei 5,385, was knapp über dem Wert von Zeitpunkt 1 liegt. Typische Beispiele für MSE zu diesem Zeitpunkt mit fünf bzw. sechs Wörtern finden sich in den Belegen [11]–[14]:

[11] *Dieses Fitnessstudio mag ich besonders*, […] (TN004; TP3; fünf Wörter bzw. Token)

[12] *Deshalb hebe ich die Gewichte.* (TN013; TP3, sechs Wörter bzw. Token)

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

[13] *Wenn du nächste woche Zeit hast,* [...] (TN029; TP3; sechs Wörter bzw. Token)

[14] *ich muss rund 20 Kilo abnehmen* (TN020; TP3; sechs Wörter bzw. Token)

Der Wert der Subordinationsrate ist gleichzeitig im Vergleich zu Zeitpunkt 2 um etwa 70 % von 0,105 auf 0,177 gestiegen. Fast jede fünfte satzwertige Einheit beinhaltet zu diesem Zeitpunkt ein Verb in Endstellung in einem subordinierenden Nebensatz. Die geschilderten Veränderungen könnten ein Hinweis darauf sein, dass die Lernenden zu diesem Zeitpunkt zwar kürzere, aber dafür ‚komplexere' Satzstrukturen produzieren. Der Mittelwert für die Rate komplexer Nominalphrasen ist mit 0,322 im Vergleich zu Zeitpunkt 2 fast gleichgeblieben. Auffällig ist, dass die Standardabweichungen für die Variablen ‚Wörter' und ‚Satzwertige Einheiten' etwas größer geworden sind, die Teilnehmenden sich folglich in der Textlänge etwas deutlicher unterscheiden als noch zu Zeitpunkt 2. Für die Variablen ‚MLC' und ‚DC/C' ist sie etwas kleiner, während sie für ‚ComNP/C' im Vergleich zu Zeitpunkt 2 minimal zunimmt.

4.1.2 Ergebnisse: Emergenz- und Profilanalysen

In diesem Abschnitt werden die Ergebnisse der Emergenz- bzw. Profilanalysen nach Grießhaber, die für alle Lernertexte zu den drei Zeitpunkten durchgeführt werden, vorgestellt. Dabei werden zunächst deskriptive Statistiken zu den produzierten Wortstellungsmustern und den erreichten Profilstufen präsentiert. Eine Profilstufe gilt in einem Text als erreicht, wenn die jeweilige Struktur mindestens dreimal korrekt in einem Text produziert wird (vgl. zur Technik der Profilanalyse die Kapitel 2.4.3 und 3.2.4). In einem zweiten Schritt innerhalb dieses Abschnitts werden die Ergebnisse des Friedman-Tests auf Gruppenebene gezeigt, der an dieser Stelle testet, ob sich die ausgewählte Teilstichprobe der 44 Lernenden hinsichtlich der erreichten Profilstufe über die drei untersuchten Zeitpunkte signifikant unterscheidet. Daran anschließend wird die Emergenz verschiedener sprachlicher Phänomene anhand von zwei verschiedenen Emergenzkriterien in Bezug auf die drei ausgewählten Zeitpunkte in den Blick genommen und überprüft, ob die Erwerbsstufen in der vorausgesagten Sequenz emergent werden.

Deskriptive Statistik der Profilanalyse zu Zeitpunkt 1:

Zu Zeitpunkt 1 werden insgesamt 444 minimale satzwertige Einheiten produziert (vgl. Tabelle 15). In Tabelle 18 ist ersichtlich, wie sich diese 444 satzwertigen Einheiten auf die einzelnen Wortstellungsmuster bzw. die syntaktischen Profilstufen verteilen. 99 der produzierten Einheiten waren Fragmente oder Chunks ohne Prädikate. Dies entspricht einem prozentualen Anteil von etwa 22,3 % (vgl. Belege [15] und [16].[196]

196 Die prozentualen Anteile werden hier und im Folgenden jeweils auf die erste Nachkommastelle gerundet angegeben.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

[15] *Dickes Bussi* TN038 (TN038, TP1)

[16] *die un in Leipzig groß und altern un in Deutschland.* (TN069, TP1)

143 Einheiten werden auf Profilstufe 1 ‚Finitum' eingeordnet (~32,2 %), was bedeutet, dass sie dem Muster Subjekt-Verb-(Objekt) entsprechen (vgl. Beleg [17].

[17] *Leipzig hat sehr schön Sehenswürdigkeit,* (TN048, TP1)

90 der 444 satzwertigen Einheiten (~20,3 %) werden auf Profilstufe 2 eingeordnet und beinhalten eine Separation der beiden Prädikatsteile (vgl. Beleg [18].

[18] *Und du müss der Sehenswürdigkeit ansehen.* (TN062, TP1)

Auf die Profilstufe 3 ‚Inversion' werden zu diesem Zeitpunkt 88 der satzwertigen Einheiten eingeordnet (~19,8 %) und 24 Einheiten (~5,4 %) enthalten ein finites Verb in Endstellung (Stufe 4 ‚Verbendstellung') (vgl. Belege [19] und [20]).

[19] *In diese Wochenende werde Ich mit den Freunden ein Picknick machen.* (TN061, TP1)

[20] Das ist meine Planen, **wenn du Zeit hast.** (TN064, TP1)

Auf den Profilstufen 5 und 6 werden keine Strukturen produziert.

Deskriptive Statistiken zu den produzierten Wortstellungsmustern – TP1								
		Stufe 0 Fragmente / Chunks	Stufe 1 Finitum	Stufe 2 Separation	Stufe 3 Inversion	Stufe 4 Verbend	Stufe 5 Integrierter NS	Stufe 6 EPA
N	Gültig	44	44	44	44	44	44	44
	Fehlend	0	0	0	0	0	0	0
Mittelwert		2,25	3,25	2,05	2,00	0,55	0,00	0,00
Standardfehler des Mittelwerts		0,146	0,252	0,206	0,225	0,119	0,000	0,000
Median		2,00	3,00	2,00	2,00	0,00	0,00	0,00
Modus		2	3	2	2	0	0	0
Std.-Abweichung		0,967	1,672	1,363	1,494	0,791	0,000	0,000
Varianz		0,936	2,797	1,858	2,233	0,626	0,000	0,000
Minimum		1	0	0	0	0	0	0
Maximum		6	7	5	6	3	0	0
Summe		99	143	90	88	24	0	0

Tabelle 18: Deskriptive Statistiken für die zu TP1 produzierten Wortstellungsmuster

Anhand der ermittelten Profilstufen für die einzelnen satzwertigen Einheiten werden zwei der 44 Lernertexte zu TP1 auf Profilstufe 0 ‚Bruchstücke', 13 Lernertexte auf Profilstufe 1 ‚Finitum', 13 Lernertexte auf Profilstufe 2 ‚Separation' und 15 Lernertexte auf Profilstufe 3 ‚Inversion' eingeschätzt (vgl. Tabelle 18). Etwas mehr als ein Drittel der Lernenden der Stichprobe erreichen zu diesem Zeitpunkt Profilstufe 3. Für einen Lernertext wird zu diesem Zeitpunkt eine Ein-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

stufung auf Profilstufe 4 ‚Verbendstellung' vorgenommen. Die Daten finden sich zusammengefasst und ergänzt mit den prozentualen Anteilen der einzelnen erreichten Profilstufen in Tabelle 19.

TP1_Erreichte Profilstufen					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	0	2	4,5	4,5	4,5
	1	13	29,5	29,5	34,1
	2	13	29,5	29,5	63,6
	3	15	34,1	34,1	97,7
	4	1	2,3	2,3	100,0
	Gesamt	44	100,0	100,0	

Tabelle 19: Häufigkeitsübersicht über die zu TP1 erreichten Profilstufen

Deskriptive Statistik der Profilanalyse zu Zeitpunkt 2:

Zu Zeitpunkt 2 werden 573 satzwertige Einheiten produziert (vgl. Tabelle 16). 54 der 573 (~9,4 %) satzwertigen Einheiten sind Bruchstücke auf Stufe 0 (vgl. Tabelle 20; vgl. Beleg [20]).

[20] *Aber das wetter in Berlin schlesstes wetter.* (TN016, TP2)

Die meisten der satzwertigen Einheiten, insgesamt 209, werden auf Stufe 1 ‚Finitum' eingeordnet, was einem prozentualen Anteil von etwa 36,5 % entspricht (vgl. Beleg [21].

[21] *die Temparatur legt bie –2 Grad.* (TN005, TP2)

127 (~22,2 %) satzwertige Einheiten entsprechen Stufe 2 ‚Separation' (vgl. Beleg [22].

[22] *Ich habe mit ihm die stadt besichtgt.* (TN028, TP2)

125 (~21,8) satzwertige Einheiten werden der Stufe 3 ‚Inversion' zugeordnet (vgl. Beleg [23].

[23] *Zuerst haber wir einen Rundgang über die Stadt gemacht.* (TN033, TP2)

Es werden mehr als doppelt so viele satzwertige Einheiten auf Stufe 4 ‚Verbendstellung', nämlich 55, produziert. Dies entspricht einem prozentualen Anteil von etwa 9,6 % an allen produzierten satzwertigen Einheiten (vgl. Beleg [24]).

[24] *dann haben wir ins Restaurant gegessen,* ***weil wir sehr Hungirig sind.*** (TN031, TP2)

Zum ersten Mal werden außerdem zwei Strukturen (~0,35 %) auf Profilstufe 5 ‚Integrierter Nebensatz' produziert (vgl. Beleg [25]).

[25] *und ich würde die Leute, die ich getroffen habe und euch unterhalten haben, wieder sehen!* (TN004, TP2)

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Deskriptive Statistiken zu den produzierten Wortstellungsmustern – TP2								
		Stufe 0 Fragmente / Chunks	Stufe 1 Finitum	Stufe 2 Separation	Stufe 3 Inversion	Stufe 4 Verbend	Stufe 5 Integrierter NS	Stufe 6 EPA
N	Gültig	44	44	44	44	44	44	44
	Fehlend	0	0	0	0	0	0	0
Mittelwert		1,23	4,75	2,89	2,84	1,25	0,05	0,00
Standardfehler des Mittelwerts		0,134	0,309	0,265	0,313	0,214	0,032	0,000
Median		1,00	5,00	3,00	3,00	1,00	0,00	0,00
Modus		1	4[a]	1	1[a]	0	0	0
Std.-Abweichung		0,886	2,047	1,755	2,079	1,416	0,211	0,000
Varianz		0,784	4,192	3,080	4,323	2,006	0,044	0,000
Minimum		0	0	0	0	0	0	0
Maximum		5	9	6	8	5	1	0
Summe		54	209	127	125	55	2	0
a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.								

Tabelle 20: Deskriptive Statistiken für die zu TP2 produzierten Wortstellungsmuster

Zu Zeitpunkt 2 wird keiner der Texte mehr auf Profilstufe 0 eingestuft und nur vier der Texte auf Profilstufe 1 ‚Finitum'. Die meisten der Texte werden auf den Profilstufen 2 (14 Texte) und 3 (17 Texte) eingeordnet. Neun der 44 Texte befinden sich auf Profilstufe 4 ‚Verbendstellung'. Es fällt hier eine Verschiebung zu den höheren Profilstufen im Vergleich zu Zeitpunkt 1 auf. Der Anteil an Texten auf den Stufen 3 und 4 nimmt deutlich zu. Wie zu Zeitpunkt 1 werden zu Zeitpunkt 2 keine Texte den Stufen 5 oder 6 zugeordnet. Die Daten und die prozentualen Anteile der erreichten Profilstufen finden sich zusammengefasst in Tabelle 21.

TP2_Erreichte Profilstufen					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1	4	9,1	9,1	9,1
	2	14	31,8	31,8	40,9
	3	17	38,6	38,6	79,5
	4	9	20,5	20,5	100,0
	Gesamt	44	100,0	100,0	

Tabelle 21: Häufigkeitsübersicht über die zu TP2 erreichten Profilstufen

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Deskriptive Statistik der Profilanalyse zu Zeitpunkt 3:

Zu Zeitpunkt 3 werden 846 satzwertige Einheiten produziert (vgl. Tabelle 17: Deskriptive Daten zu den produzierten Texten zu Zeitpunkt 3 (TP3)). 135 von 846 satzwertigen Einheiten werden als Bruchstücke auf Stufe 0 klassifiziert (vgl. Tabelle 22; vgl. Beleg [26]). Dies entspricht einem prozentualen Anteil von fast 16 %.

[26] *Meine neue Nachrichtung,* (TN001, TP3)

Wie schon zu Zeitpunkt 2 werden die meisten der produzierten Einheiten der Stufe 1 ‚Finitum' zugeordnet. Insgesamt ist dies für 292 (~34,5 %) Einheiten der Fall (vgl. Beleg [27].

[27] *Ich laufe gern auf die Maschine,* (TN013, TP3)

109 satzwertige Einheiten (~12,9 %) entsprechen Profilstufe 2 ‚Separation' und 181 der Einheiten (~21,4 %) der Profilstufe 3 ‚Inversion' (vgl. Belege [28] und [29].

[28] *Und mein Arzt hat mir gesagt* (TN027, TP3)

[29] *Dort habe ich viele neuen Freunde getroffen* (TN029, TP3)

Als Stufe 4 ‚Verbendstellung' werden 129 satzwertige Einheiten (1~15,2 %) klassifiziert (vgl. Beleg [30]). Eine Einheit wird auf Stufe 5 ‚Integrierte NS' produziert (~0,1 %) (vgl. Beleg [31]).

[30] *Ich habe besonders Laufmachine triniert,* ***weil ich abnehmen möchte.*** (TN028, TP3)

[31] *natülich, wenn Du Zeit hast und etwas erleben möchtest kann ich über den Fest weider erzählen* (TN064, TP3)

Es werden zu Zeitpunkt 3 von den Lernenden der Stichprobe keine Strukturen auf Profilstufe 6 ‚EPA' produziert.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Deskriptive Statistiken zu den produzierten Wortstellungsmustern – TP3								
		Stufe 0 Fragmente / Chunks	Stufe 1 Finitum	Stufe 2 Separation	Stufe 3 Inversion	Stufe 4 Verbend	Stufe 5 Integrierter NS	Stufe 6 EPA
N	Gültig	44	44	44	44	44	44	44
	Fehlend	0	0	0	0	0	0	0
Mittelwert		3,07	6,64	2,48	4,11	2,93	0,02	0,00
Standardfehler des Mittelwerts		0,173	0,467	0,273	0,343	0,313	0,023	0,000
Median		3,00	6,00	2,00	3,50	3,00	0,00	0,00
Modus		3	5[a]	1	3	3	0	0
Std.-Abweichung		1,149	3,096	1,811	2,274	2,073	0,151	0,000
Varianz		1,321	9,586	3,279	5,173	4,298	0,023	0,000
Minimum		0	2	0	0	0	0	0
Maximum		6	15	8	10	9	1	0
Summe		135	292	109	181	129	1	0
a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.								

Tabelle 22: Deskriptive Statistiken für die zu TP3 produzierten Wortstellungsmuster

Wie schon zu Zeitpunkt 2 werden zu Zeitpunkt 3 keine Texte der Profilstufe 0 mehr zugeordnet. Jeweils nur zwei Texte befinden sich auf den Stufen 1 ‚Finitum' und 2 ‚Separation'. Mehr als 90 % der Texte werden nun den Profilstufen 3 und 4 zugerechnet, darunter 15 der Texte der Profilstufe 3 ‚Inversion' und 25 der Texte der Profilstufe 4 ‚Verbendstellung'. Wie schon zu den beiden vorangegangenen Zeitpunkten werden keine Texte der Stichprobe den Profilstufen 5 oder 6 zugeordnet. Tabelle 23 gibt eine Übersicht über die zu TP3 erreichten Profilstufen und die entsprechenden prozentualen Anteile.

TP3_Erreichte Profilstufen					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1	2	4,5	4,5	4,5
	2	2	4,5	4,5	9,1
	3	15	34,1	34,1	43,2
	4	25	56,8	56,8	100,0
	Gesamt	44	100,0	100,0	

Tabelle 23: Häufigkeitsübersicht über die zu TP3 erreichten Profilstufen

Zusammenfassend zeigen die deskriptiven Daten der Profilanalysen zu den drei Zeitpunkten, dass relativ früh in der (gesteuerten) zweitsprachlichen Entwicklung (hier TP1) zahlreiche ‚komplexe' Strukturen der Profilstufen 3 und 4 pro-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

duziert werden und dies in einem solchem Umfang, dass die Texte als Ganzes diesen Profilstufen zugeordnet werden. Zu Zeitpunkt 2 machen Texte auf diesen beiden Profilstufen deutlich mehr als die Hälfte der Texte aus. Erwartungsgemäß werden zunächst weniger ‚Bruchstücke' produziert. Der Anteil an Bruchstücken nimmt zwischen TP2 und TP3 wieder zu. Diese Zunahme lässt sich zu einem großen Teil durch ‚beabsichtigte' Strukturen wie Gruß- oder Abschiedsformeln und nicht etwa durch ungrammatische satzwertige Einheiten erklären. Zu Zeitpunkt 3 werden zwar nur etwas weniger als 37 % Prozent der satzwertigen Einheiten den Profilstufen 3 und 4 zugeordnet. Diese sind so über die Texte verteilt, dass mehr als 90 % der Texte auf diesen beiden Profilstufen eingeordnet werden. Dass die Streuung innerhalb der Stichprobe mit jedem Zeitpunkt abnimmt und zu Zeitpunkt 3 am geringsten ist, verdeutlicht Abbildung 17:

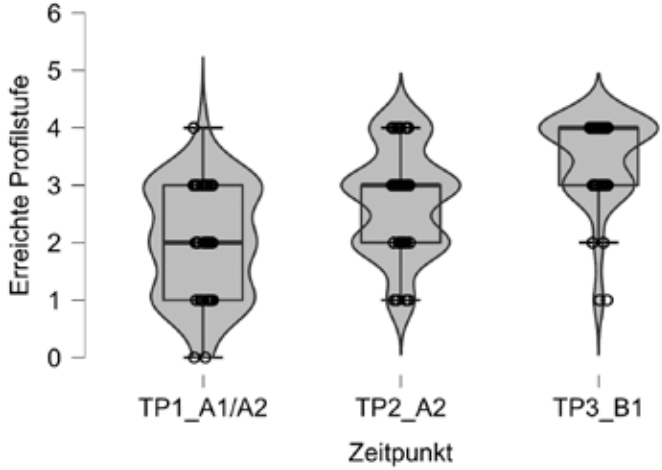


Abbildung 17: Boxplot mit Violinenelementen zu den erreichten Profilstufen zu TP1, TP2 und TP3. Gezeigt werden die Daten für dieTeilstichprobe (n=44).[197]

In Abbildung 17 wird die Streuung der Daten der Teilnehmenden für die drei Zeitpunkte mithilfe eines Boxplots mit zusätzlichen Violinen-Elementen dargestellt. Auf diesem ist der Interquartilsabstand (Box mit unterem, oberem und mittlerem Quartil oder Median) zu sehen. Die etwas dickere schwarze Linie in der Mitte zeigt jeweils den Median oder das 50 %-Perzentil. Die Antennen (*Whisker*) erstrecken sich auf das 1,5-fache des Interquartilsabstandes und helfen dabei, Ausreißer (zur Definition von Ausreißern in der vorliegenden Studie vgl. Fußnote 161) in den Daten schneller zu identifizieren. So liegt etwa zu TP3 ein Datenpunkt deutlich außerhalb des durch die Antennen gekennzeichneten Bereichs. Zusätzlich sind die Boxplots als Violinen-Elemente dargestellt, die die Streuung und die Vertei-

197 Diese Grafik wurde mit JASP in der Version 0.14.1 erstellt (JASP Team 2020).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

lungsdichte der Daten detaillierter verdeutlichen als herkömmliche Boxplots. Der Umfang der Violinenbäuche zeigt an, wo sich viele (‚dickerer Bauch') und wo sich weniger (‚kleinerer Bauch') Daten befinden. Während zu Zeitpunkt 1 noch eine relativ gleichmäßige Verteilung über die Profilstufen 0 bis 3 mit nur einem Datenpunkt auf Profilstufe 4 zu beobachten ist, verschiebt sich die Verteilung über die nächsten beiden Zeitpunkte immer mehr in Richtung der beiden oberen Profilstufen 3 und 4. Zu Zeitpunkt 3 zeigt sich die deutlichste Ausprägung des Violinenbauches auf der Profilstufe 4, während sich nur noch sehr wenige Datenpunkte auf den unteren Profilstufen finden. Ausreißer werden in Abbildung 17 mit nicht gefüllten und fettgedruckten Kreisen angezeigt. Die Boxplots deuten darüber hinaus auf einen linearen bzw. kontinuierlichen Lernverlauf hin, bei dem die Varianz innerhalb der Gruppe von TP1 zu TP3 konstant abnimmt, während gleichzeitig auf Gruppenebene immer höhere syntaktische Profilstufen erreicht werden (TP1: *Mdn* = 2; TP2: *Mdn* = 3, TP3: *Mdn* = 4).

Deskriptive Statistiken zu den erreichten Profilstufen – TP1, TP2 und TP3				
		TP1_A1/A2	TP2_A2	TP3_B1
N	Gültig	44	44	44
	Fehlend	0	0	0
Mittelwert		2,00	2,70	3,43
Standardfehler des Mittelwerts		0,145	0,136	0,119
Median		2,00	3,00	4,00
Modus		3	3	4
Std.-Abweichung		0,964	0,904	0,789
Varianz		0,930	0,818	0,623
Minimum		0	1	1
Maximum		4	4	4
Summe		88	119	151
Perzentile	25	1,00	2,00	3,00
	50	2,00	3,00	4,00
	75	3,00	3,00	4,00

Tabelle 24: Deskriptive Daten zum Friedman-Test hinsichtlich der erreichten syntaktischen Profilstufe

Um zu prüfen, ob sich die Daten bezüglich der erreichten Profilstufen über die Zeit verändern, wird ein Friedman-Test durchgeführt. Dieser wird hier als nicht-parametrische Alternative zu einer Varianzanalyse mit Messwiederholung (*repeated measures ANOVA*) durchgeführt, da die Voraussetzungen der Normalverteilung und der Varianzhomogenität innerhalb der Stichprobe nicht erfüllt sind. Der Friedman-Test ist signifikant und zeigt, dass sich das syntaktische Niveau der Teilnehmenden über die drei Zeitpunkte signifikant verändert, $\chi^2(2)=37{,}468$, $p<0{,}000$, Kendall's W = 0,378 (vgl. Tabelle 25).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Zusammenfassung der zweifaktoriellen Varianzanalyse für Ränge nach Friedman bei verbundenen Stichproben	
Gesamtzahl	44
Teststatistik	37,468
Freiheitsgrad	2
Asymptotische Sig. (zweiseitiger Test)	0,000

Tabelle 25: Zusammenfassung der Ergebnisse für den Friedman-Test

Paarweise Vergleiche mit Dunn-Bonferroni-Korrektur für mehrfache Vergleiche zeigen folgende Unterschiede zwischen den Zeitpunkten:

– Eine signifikante Veränderung mit kleiner Effektgröße (Pearson's r; vgl. Field 2018: 116f.) wird von TP1 zu TP2, $p = 0{,}032$, T=–0,545, $r = -0{,}27$ beobachtet.[198]
  Eine signifikante Veränderung mit großer Effektgröße wird von TP1 zu TP3, $p > 0{,}000$, T= –1,159, $r = -0{,}58$ beobachtet.
  Eine signifikante Veränderung mit mittlerem Effekt wird von TP2 zu TP3, $p = 0{,}012$, T= –0,614, $r = -0{,}31$ beobachtet.

Deutlich wird vor allem der signifikante Unterschied mit großer Effektstärke zwischen Zeitpunkt 1 und Zeitpunkt 3.[199] Dieser Vergleich deckt das größte zeitliche Intervall ab, was Unterschiede innerhalb eines voranschreitenden Lernprozesses sehr wahrscheinlich macht. Die Unterschiede, die sich jeweils nur auf das Intervall zwischen einem Zeitpunkt und dem direkt darauffolgenden beziehen, sind moderater. Zwischen Zeitpunkt 1 und Zeitpunkt 2 zeigen sich die geringsten Unterschiede. Dies stimmt mit den Tendenzen überein, die in den in Abbildung 17 dargestellten Verteilungen und in den deskriptiven Daten deutlich wurden. Gerade zu Zeitpunkt 3 kommt es zu einer deutlichen Verschiebung

198 Die Effektgrößen (Pearson's *r*) werden berechnet, indem die in z-Scores standardisierte Teststatistik durch die Quadratwurzel aus 88 geteilt wird. 88, da zwei Gruppen mit jeweils 44 Teilnehmenden verglichen werden. Dabei ist es nicht relevant, dass die Beobachtungen von denselben Teilnehmenden kommen (vgl. Field 2018: 328).

199 Die Lernenden der Teilstichprobe besuchten innerhalb der Studie verschiedenen Kursgruppen, die jeweils zu unterschiedlichen Zeitpunkten starteten (vgl. Kapitel 3.1.3). Daher werden im Folgenden durchschnittliche Zeitangaben für die drei Zeitpunkte angegeben (die Angaben zu den Medianen und den Standardabweichungen finden sich in Fußnote 154). Zeitpunkt 1 war für die Lernenden der Teilstichprobe im Schnitt nach 15,23 Wochen. Die Lernenden hatten zu diesem Zeitpunkt im Schnitt 244,52 Unterrichtseinheiten besucht. Zeitpunkt 2 war für die Teilnehmenden im Schnitt nach 25,14 Wochen und 390,00 besuchten UE. Zeitpunkt 3 war nach durchschnittlich 29,43 Wochen und 485,36 besuchten UE. Zwischen Zeitpunkt 1 und Zeitpunkt 2 lagen etwa neun Wochen und zwischen Zeitpunkt 2 und 3 knapp vier Wochen. Der zeitliche Abstand zwischen Zeitpunkt 1 und Zeitpunkt 3 betrug ca. 13 Wochen.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

hin zu den höheren Profilstufen 3 und 4 in der jeweiligen Gesamtbewertung der Lernertexte.

Von diesen Ergebnissen ausgehend werden die Ergebnisse zweier (longitudinaler) Emergenzanalysen vorgestellt, die einerseits Grießhabers (2012) Kriterium von drei Vorkommen einer Struktur und andererseits Pienemanns (1998) Operationalisierung des Emergenzkriteriums von mindestens einem korrekt realisierten Vorkommen in mindestens vier obligatorischen Kontexten zur Grundlage haben (vgl. zum Vorgehen in diesem Abschnitt Kapitel 3.2.4). Hinsichtlich des Emergenzkriteriums nach Pienemann (1998) wird zusätzlich nach Belegen dafür gesucht, dass eine bestimmte Struktur noch nicht erworben wurde. Pienemann (1998) sieht dies dann als gegeben an, wenn sich vier obligatorische Kontexte innerhalb eines Textes finden, in denen die Zielstruktur nicht produziert wird.

Wird zunächst Grießhabers Kriterium an die Daten der drei Zeitpunkte angelegt, ergibt sich kein eindeutiges Bild, was die sequenziell-implikationelle Abfolge der Erwerbsstufen angeht. Wenn nur das erstmalige Auftreten einer Struktur gemäß des Emergenzkriteriums und die sich daraus ergebende Reihenfolge in Betracht gezogen werden, folgen 30 der 44 Lernenden nicht den angenommenen Sequenzen. Nur für ein Drittel der Teilnehmenden lässt sich die vorhergesagte Sequenz auf Grundlage der Emergenz der Strukturen rekonstruieren. Zu Zeitpunkt 1 erreichen 19 Lernende durch dreimaliges Auftreten der jeweiligen Strukturen höhere Profilstufen, während niedrige Profilstufen gleichzeitig nicht erreicht werden bzw. sich nicht drei korrekte Realisierungen für die Strukturen der niedrigeren Profilstufen finden lassen. Zu Zeitpunkt 2 ist dies für 20 und zu Zeitpunkt 3 für 27 Lernende der Fall.[200]

Pienemann definiert Emergenz und damit das Erreichen einer Erwerbsstufe durch das einmalige Auftreten einer korrekt realisierten Struktur beim gleichzeitigen Vorkommen von mindestens vier obligatorischen Kontexten, in denen diese Struktur produziert werden müsste.[201] Wird dieses Kriterium auf die drei Zeitpunkte angewendet und jeweils nur das erstmalige Auftreten einer Struktur über die drei Zeitpunkte in den Blick genommen, folgen nur neun (20,5 %) der 44 Teilnehmenden den vorausgesagten Sequenzen**.** Für 35 (79,5 %) der Lernenden lässt sich dies nicht belegen. Zu Zeitpunkt 1 zeigt sich unter Anwendung des Emergenzkriteriums, dass 18 der 44 Lernerinnen und Lerner eine höhere Erwerbsstufe erreichen, während sich gleichzeitig nicht genügend Belege für das

200 Dies ist im Kontext der PT und der PA zwar kein Widerspruch, sollte aber im Blick behalten werden, da dieser Umstand verdeutlicht, wie stark diagnostische Verfahren davon abhängig sind, dass genügend obligatorische Kontext für einzelne Strukturen elizitiert werden müssten (vgl. zu diesem Punkt Kapitel 5).

201 Zur Erinnerung: Als Beleg dafür, dass eine Erwerbsstufe noch nicht erreicht wurde, gilt das Vorkommen von mindestens vier obligatorischen Kontexten, ohne dass eine einzige Struktur korrekt realisiert wird.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Erreichen der niedrigeren Stufen finden lassen. Das könnte bedeuten, dass in einem der lernersprachlichen Texte weniger als vier obligatorische Kontexte einer Struktur gefunden werden und der Text daher über das Emergenzkriterium nach Pienemann (1998) keine Information über die Emergenz einer Struktur liefern kann bzw. aus der Analyse ausgeschlossen wird. Ebenfalls könnte es bedeuten, dass in den Texten eine Erwerbsstufe tatsächlich noch nicht erreicht wurde (vgl. dazu Kapitel 5). Ob dies allerdings der Fall ist, lässt sich letztlich nicht entscheiden bzw. nachweisen, da das Fehlen von Belegen nicht als das Nichtvorhandensein einer Struktur gewertet werden kann. Von den 44 Lernenden erreichen mit diesem Emergenzkriterium neun Lernende Stufe 1 ‚Finitum' als höchste zu diesem Zeitpunkt emergierte Erwerbsstufe, neun Lernende Stufe 2 ‚Separation' und elf Lernende Stufe 3 ‚Inversion'. Für 15 der Lernenden kann keine Emergenz einer Stufe ermittelt werden, da es sich um Texte mit nur sehr wenigen MSE handelt, in denen für keine der Erwerbsstufen ausreichend obligatorische Kontexte vorliegen.

Zu Zeitpunkt 2 etwa finden sich bei 26 der Lernenden Belege dafür, dass eine höhere Erwerbsstufe emergiert ist, bei gleichzeitigen fehlenden Belegen darüber, dass die niedrigeren Stufen emergiert sind. Sieben der Lernenden erreichen mit diesem Emergenzkriterium zu Zeitpunkt 2 Stufe 1 als höchste zu diesem Zeitpunkt emergierte Stufe, neun Lernende Stufe 2, 21 Lernende Stufe 3 und vier Lernende erreichen Stufe 4. Für drei Teilnehmende kann aufgrund fehlender obligatorischer Kontexte keine Zuordnung vorgenommen werden.

Zum dritten Zeitpunkt erreichen 31 der 44 Lernenden eine höherliegende Erwerbsstufe, ohne dass sich in den Texten Belege für die Emergenz darunterliegender Erwerbsstufen finden. Zu diesem Zeitpunkt ist für fünf Lernende Stufe 1 die höchste im Text emergente Erwerbsstufe, für drei Lernende Stufe 2, für 17 Lernende Stufe 3 und für 19 Lernende Stufe 4.

Belege dafür, dass zu einzelnen Zeitpunkten Erwerbsstufen noch nicht erreicht sind, können anhand des Kriteriums von vier obligatorischen Kontexten bei gleichzeitiger Nichtrealisierung der Struktur in diesen Kontexten nur in zwei Fällen für die gesamte Stichprobe über alle drei Zeitpunkte gefunden werden. TN00X produziert zu Zeitpunkt 2 fünf obligatorische Inversionskontexte, in denen in keinem Fall eine Inversion realisiert wird. TN0X0 produziert zu Zeitpunkt 3 insgesamt vier obligatorische Kontexte für eine Verbendstellung, von denen keine realisiert wird. In einigen Fällen werden zwar Strukturen in vier obligatorischen Kontexten innerhalb eines Textes nicht produziert, während es gleichzeitig zu ‚korrekt' realisierten Varianten kommt: TN038 produziert zu Zeitpunkt 1 insgesamt sechs obligatorische Kontexte für die Stufe ‚Separation', in denen in zwei Fällen korrekte Strukturen produziert werden. Diese sehr geringe Anzahl an Belegen für die Nichtemergenz von Strukturen deutet unter Umständen darauf hin, dass dieses Kriterium mit sehr großer Vorsicht eingesetzt werden muss (für eine Diskussion der Emergenzkriterien im Kontext der Ergebnisse dieser Untersuchung vgl. Kapitel 5). Tabelle 26 stellt die beiden verwendeten

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Emergenzkriterien gegenüber und zeigt Gemeinsamkeiten und Unterschiede auf. Es wird deutlich, dass sich der Blick auf die gesamte Gruppe, aber ebenfalls auf einzelne Lernende (und damit indirekt auf zweitsprachliche Erwerbs- bzw. Entwicklungsverläufe) deutlich verändert, je nachdem, welches Emergenzkriterium angelegt bzw. wie Emergenz operationalisiert wird. Das Emergenzkriterium nach Grießhaber sorgt einerseits dafür, dass zu jedem Zeitpunkt alle lernersprachlichen Texte einer Profilstufe zugeordnet werden. Andererseits werden die Lernertexte im Vergleich zum Kriterium nach Pienemann tendenziell auf höheren Profilstufen eingestuft. Auffällig für die Daten, die nach Pienemanns Emergenzkriterium eingestuft werden, ist neben der angesprochenen tendenziell etwas niedrigeren Einordnung der Texte vor allem der Umstand, dass viele Texte nicht genügend Belegstellen besitzen, um sie einzuordnen. Dies ist vor allem zu Zeitpunkt 1 auffällig, zu dem 15 Texte (folglich mehr als ein Drittel) nicht genügend Belegstellen besitzen, um sie einer Erwerbsstufe zuzuordnen (vgl. Tabelle 26).[202] Ab Zeitpunkt 2 scheinen die Texte ebenfalls für das Emergenzkriterium nach Pienemann ausreichend Belegstellen aufzuweisen, denn zu diesem Zeitpunkt können nur noch drei Texte nicht eingestuft werden. Zu Zeitpunkt 3 finden sich in allen Lernertexten genügend Belegstellen, um Einstufungen nach Pienemann vorzunehmen. Es zeigt sich, dass die beiden Emergenzkriterien mit der Zeit wieder konvergieren und die Unterschiede zwischen ihnen über die drei Zeitpunkte geringer werden. Dies könnte durchaus mit einer zunehmenden Textlänge und einer damit einhergehenden größeren Anzahl an MSE pro Text zusammenhängen. Während zu Zeitpunkt 1 noch die Hälfte der Lernenden unterschiedlich eingestuft wird, sind es zu Zeitpunkt 2 nur noch 14 Lernenden und zu Zeitpunkt 3 noch 12 Lernende, folglich weniger als ein Drittel. Da Stufe 4 gleichzeitig das (vorläufige) Deckenniveau für die Lernenden der ausgewählten Stichprobe darzustellen scheint, ist eine Konvergenz zu dem Zeitpunkt, an dem viele Lernende diese Stufe erreichen, relativ erwartbar.

202 Dieser Umstand verdeutlicht eine grundlegende Herausforderung der Lernersprachanalyse. Wie lange muss ein Text sein bzw. wie viele Token oder Strukturen muss er umfassen, um verlässlich eingestuft werden zu können? Vgl. zu dieser Herausforderung die Diskussion in Kapitel 5 Diskussion. Diese Herausforderung wird bereits bei Pienemann (1998) ausführlich diskutiert (vgl. Pienemann 1998: 145–148).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

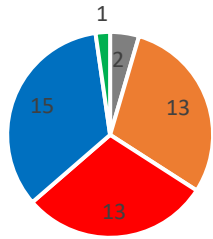
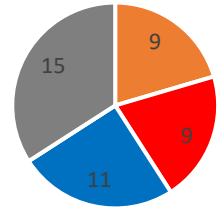
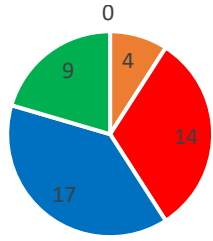
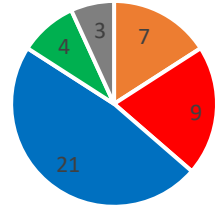
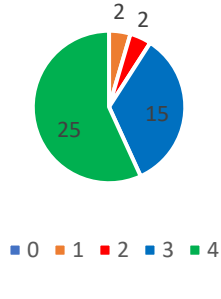
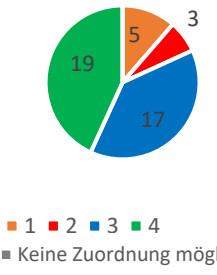
Zeitpunkt	Emergenz nach Grießhaber	Emergenz nach Pienemann	Unterschiede zwischen Lernenden
TP1			Für 22 Lernende ergeben die unterschiedlichen Emergenzkriterien dieselbe Erwerbsstufe zu diesem Zeitpunkt, für 22 Lernende ergeben die beiden Kriterien differierende Ergebnisse (15 Lernende ohne Zuordnung einbegriffen)
TP2			Für 30 Lernende ergeben sich zu TP2 dieselben Erwerbsstufen. Für 14 Lernende ergeben sich für die beiden Emergenzkriterien unterschiedliche Erwerbsstufen.
TP3			Für 32 Lernende ergeben sich zu TP3 dieselben Erwerbsstufen. Für 12 Lernende ergeben sich für die beiden Emergenzkriterien unterschiedliche Erwerbsstufen.

Tabelle 26: Vergleich der beiden Emergenzkriterien nach Grießhaber (2012) und Pienemann (1998)

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

An dieser Stelle sind zwei Dinge zu beachten: Einerseits besitzen die beiden Stufen 2 ‚Separation' und 3 ‚Inversion' eine Schnittmenge, die die Analysen hier nicht ganz präzise erscheinen lässt, da Lernende durchaus Inversionsstrukturen mit zusätzlich in diesen MSE realisierten Separationen produzieren, diese dann allerdings der Stufe ‚Inversion' zugeordnet werden.[203] Dies zeigen beispielhaft die Belege [32], [33] und [34]:

[32] ***deshalb lade*** *ich dich nach hier her* ***ein***. (TN004, TP1)

[33] ***Dann sind*** *wir an den Berlinturm* ***gestiegt***. (TN012, TP2)

[34] ***nächsten Montag würden*** *wir neus Trineng* ***machen***. (TN028, TP3)

Andererseits ist anzumerken, dass die Abwesenheit eines Beleges für die Verwendung einer Struktur keinen hinreichenden Beleg für die Abwesenheit dieser Struktur in der Lernersprache generell darstellt.[204] Fehlende bzw. nicht realisierte Strukturen könnten zufällig nicht produziert werden. Daher wird in einem zweiten Analyseschritt der longitudinale Charakter der Daten stärker miteinbezogen und Sequenzen nur dann als ‚nicht der Erwerbssequenz folgend' gewertet, wenn zu allen drei Zeitpunkten keine den Emergenzkriterien entsprechenden Belege für den Erwerb einer niedrigeren Erwerbsstufe vorliegen, gleichzeitig aber Strukturen einer höheren Profilstufe zu zwei bzw. drei Zeitpunken in ausreichender Menge für die beiden Emergenzkriterien gefunden werden. Es könnte in diesen Fällen natürlich wieder so sein, dass Strukturen zu zwei aufeinanderfolgenden Zeitpunkten zufällig nicht produziert werden. Hier werden in der gesamten Stichprobe nur wenige Fälle gefunden, in denen dies für die beiden Emergenzkriterien zutreffend ist. Für das Kriterium nach Grießhaber kann nur ein Fall identifiziert werden, bei dem durch die/den Lerner/-in zu jedem der drei Zeitpunkte ausreichend Belege für eine höhere Profilstufe gefunden werden können, während Belege für darunterliegende Stufen in den Texten zu allen Zeitpunkten fehlen. Für das Kriterium nach Grießhaber ist das für TN031 der Fall. In den Texten von TN031 können zu TP1 und TP2 nur Stufe 3 und zu TP3 nur die Stufen 3 und 4 nachgewiesen werden. TN031 produziert zu allen drei Zeitpunkten nicht genügend obligatorische Kontexte und Strukturen

203 Dennoch scheint es nicht direkt nachvollziehbar, dass die Stufe der Inversion zu so frühen Zeitpunkten frequenter verwendet wird als Strukturen der Stufe Separation. Dies legt zumindest einen zeitgleichen Erwerb dieser beiden Strukturen nahe, was aus der Perspektive der *Processability Theory* ausgeschlossen ist. Die Daten könnten zusätzlich eine generelle Vorliebe der Lernenden abbilden, die vielleicht mit dem Umstand zu tun hat, dass es sich um Lernende aus gesteuerten Lernkontexten handelt, in denen etwa schriftliche Produktionen eine prominentere Rolle spielen als in ungesteuerten Lernkontexten (vgl. hierzu die Diskussion in Kapitel 5).

204 Dieser Umstand wird in der englischsprachigen Literatur oft mit dem geflügelten Ausdruck „The absence of evidence is not the evidence of absence" (beispielhaft in Brezina 2018:19; vgl. den Problemaufriss in der Einleitung zu dieser Studie) beschrieben.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

der Stufen 1 und 2, um diese als ‚erworben' nach Grießhaber nachzuweisen. Für das Emergenzkriterium nach Pienemann lassen sich zwei Fälle nachweisen, in denen durchgehend höhere Stufen emergent werden, während nicht ausreichend Belege für niedrigere Stufen vorliegen: TN067 und TN078. Für TN067 kann der Text zu TP1 nicht eingestuft werden, zu TP2 und TP3 liegen ausschließlich ausreichend Belege für die Emergenz von Stufe 3 vor. TN078 produziert zu TP1 Belege für Stufe 2, zu TP2 nur ausreichend Belege für Stufe 3 und zu TP3 finden sich nur Belegstellen in Bezug auf die Emergenz der Stufen 2 und 3. Für TN067 liegen keine Belege der Emergenz der Stufen 1 und 2 vor und für TN078 fehlen Belegstellen für Stufe 1.

Zusammenfassend lässt sich an diesem Punkt feststellen, dass die Gruppendaten die sequentielle und lineare Abfolge der Profilstufen zunächst zu stützen scheinen und dass die Werte der beobachteten Variablen über die Zeit größer werden, was für eine zunehmende Komplexität der untersuchten Texte sprechen könnte. Die durchgeführten Emergenzanalysen anhand von zwei unterschiedlichen Emergenzkriterien differenzieren das zunächst angenommene Bild deutlich. Wird ausschließlich das erstmalige den Emergenzkriterien genügende Auftreten bestimmter Strukturen betrachtet, lassen sich daraus nur für wenige Lernende sequenzielle Abfolgen der beobachteten Strukturen rekonstruieren. Deutlich wird, dass Grießhabers Kriterium von drei Vorkommen pro Text es ermöglicht, deutlich mehr Lernertexte der untersuchten Stichprobe einzustufen, als dies mit Pienemanns Kriterium der Fall ist. Hier reichen vor allem zu Zeitpunkt 1 die vorhandenen lernersprachlichen Daten oft nicht aus, um eine Einstufung vorzunehmen. Dies könnte unter Umständen mit der ursprünglichen Anwendung dieses Kriteriums auf mündlichen Daten zusammenhängen, während Grießhabers Profilanalyse sowohl für mündliche als auch für schriftliche Daten entwickelt wurde (für eine ausführlichere Diskussion vgl. Kapitel 5). Die beiden Emergenzkriterien sorgen vor allem zu Zeitpunkt 1 für sehr unterschiedliche Einstufungen der Lernenden. Mit fortschreitender Zeit konvergieren die beiden Kriterien, was mit dem angenommenen Deckenniveau der Lernenden zusammenhängen könnte.

Für nur drei Lernende kann darüber hinaus herausgearbeitet werden, dass sich über alle drei hier betrachteten Zeitpunkte ausreichend Belege für die Emergenz höherer Stufen finden lassen, während diese gleichzeitig für niedrigere Stufen in den Texten zu allen drei Zeitpunkten fehlen. Die Daten der Emergenzanalysen deuten darüber hinaus darauf hin, dass sich sowohl eine relativ große inter- als auch intraindividuelle Variabilität findet, die im nächsten Schritt in den Fokus der Analysen rückt.

4.1.3 Ergebnisse: Longitudinale Clusteranalysen

Um ein deutlicheres Bild der Variabilität der syntaktischen Entwicklung der Lernenden zu erhalten, werden longitudinale Clusteranalysen in zwei Schritten durchgeführt. Die initiale und mit den Daten zu TP1 durchgeführte hierarchi-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

sche Clusteranalyse (HCA) zur Bestimmung der optimalen Clusteranzahl für die im folgenden Schritt durchgeführten longitudinalen k-means-Clusteranalysen (zu TP1, TP2 und TP3) deutet darauf hin, dass drei Cluster aus den Daten gebildet werden können. Dies ergibt sich aus der grafischen und der statistischen Interpretation der Ergebnisse der HCA. Zunächst wird für die grafische Interpretation das Dendrogramm der HCA untersucht (vgl. Abbildung 19). In einem Dendrogramm werden die Ergebnisse auf zwei Skalen verdeutlicht; zum einen finden sich alle in die Clusteranalyse eingegangenen Fälle (im vorliegenden Fall die Lernenden). Dies ist die senkrechte y-Achse in Abbildung 19. Zum anderen wird auf der horizontalen x-Achse die Höhe (*height*) dargestellt, an der zwei Objekte (bzw. Lernende oder Cluster von Lernenden) aufgrund ihrer Ähnlichkeit zu einem neuen Cluster zusammengefasst werden. Der Höhenwert oder Distanzwert gibt Aufschluss darüber, in welcher Reihenfolge die Fälle in Clustern zusammengefasst werden, d. h. welche Lernenden zu welchem Punkt im Clusterverfahren in ein gemeinsames Cluster eingeordnet wurden. Für alle Lernerinnen und Lerner kann so nachvollzogen werden, wann sie mit welchen anderen Lernerinnen und Lernern in ein neues Cluster zusammengefasst werden. Es ist daher möglich, einen beliebigen Höhenwert auszuwählen, um zu evaluieren, wie viele Cluster mit diesem assoziiert werden. Für den Höhenwert 5 wären das im gezeigten Dendrogramm beispielsweise fünf Cluster. Das Dendrogramm deutet darauf hin, dass drei Cluster die Stichprobe für die weitere Analyse gewinnbringend repräsentieren. Der Sprung zu zwei Clustern findet mit größerem Abstand statt als viele der Sprünge davor (Höhe = 25). Außerdem scheinen vier Cluster (Höhe = 7) im Vergleich zu einer 5-Clusterlösung (Höhe = 5) nicht viel an zusätzlichen Informationen bereitzustellen, da der Höhenabstand zwischen beiden Möglichkeiten sehr gering ist. Der Höhenunterschied zu einer 3-Clusterlösung (Höhe = 16) ist allerdings sehr deutlich. Das Dendrogramm sollte nicht allein als Maßstab dafür herangezogen werden, welche Clusteranzahl in den Daten wahrscheinlich ist.

Deshalb werden zusätzlich die Daten der Zuordnungsübersicht (*agglomeration matrix*) ausgelesen.[205] Hier werden die letzten zehn Cluster der Zuordnungsübersicht und ihre jeweiligen Fusionierungskoeffizienten für eine Interpretation berücksichtigt und in eine Visualisierung überführt, da in diesen die größten Unterschiede in den Fusionierungskoeffizienten zu erwarten sind (vgl. Abbildung 18 und generell zu diesem Vorgehen Staples / Biber 2015: 256–257 und Backhaus et al. 2018: 476–477). In dieser Abbildung wird dann nach dem Punkt gesucht, an dem die Kurve sichtbar abflacht, da ab diesem Punkt durch weitere

205 Die Zuordnungsübersicht gibt in tabellarischer Form an, bei welchem Fusionierungsschritt „welche Fälle bzw. Cluster […] durch das Ward-Verfahren bei welchem Heterogenitätsmaß zusammengefasst wurden" (Backhaus et al. 2018: 476). Außerdem wird die Fehlerquadratsumme der Fusionierungskoeffizienten angegeben, die dann zur Erstellung von Abbildung 18 genutzt werden. Die hier zugrundeliegende Zuordnungsübersicht findet sich im Anhang.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Cluster keine wesentliche, zusätzliche Information bereitgestellt wird (vgl. Staples / Biber 2015: 256). Diese Abflachung der Kurve wird „Ellenbogen-Kriterium" oder „*Elbow*-Kriterium" (Backhaus et al. 2018: 476) genannt.[206]

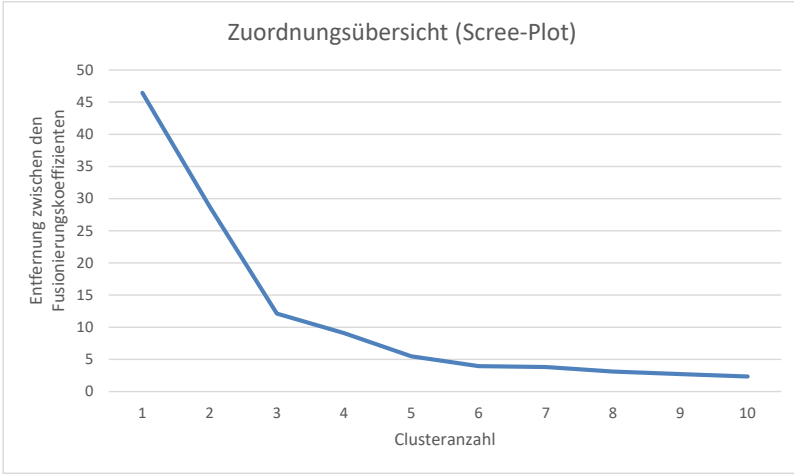


Abbildung 18: Zuordnungsübersicht zur Bestimmung der Clusteranzahl in der hierarchischen Clusteranalyse

Durch die grafische Transformation der Daten aus der Zuordnungsübersicht wird deutlich, dass die Kurve nach drei gebildeten Clustern deutlich abflacht (vgl. Abbildung 18), was darauf hindeutet, dass bei einer größeren Clusteranzahl nur noch unwesentlich mehr Informationen gewonnen werden.

Zusätzlich werden nach Staples / Biber (2015: 261–267) die Mittelwerte der Clustervariablen in Beziehung zueinander gesetzt, wenn mit einer 2- / 3- oder einer 4-Clusterlösung gearbeitet wird.[207] Diese Analyse wird mithilfe einer einfaktoriellen Varianzanalyse (*one-way ANOVA*) durchgeführt, um die Mittelwerte der untersuchten Variablen zu vergleichen (für die Ergebnisse der 3-Clusterlösung vgl. Tabelle 28). Tabelle 27 zeigt zunächst die deskriptiven Statistiken der Daten

206 Backhaus et al. (2018) machen auf den subjektiven Charakter des Elbow-Kriteriums aufmerksam und besprechen einige „Stopping-Rules" (ebd.: 477–478), die von Wissenschaftlerinnen und Wissenschaftlern zusätzlich angewendet werden können, um die Auswahl der optimalen Clusteranzahl zu begründen. Viele dieser Regeln legen eine Methodentriangulation nahe. In der vorliegenden Analyse wurden diese weiteren Ansatzpunkte nicht genutzt, da bereits eine Kombination aus verschiedenen Analysen zugrunde gelegt wurde.

207 Clustervariablen sind diese Variablen, anhand deren die Clusteranalysen durchgeführt werden. In der vorliegenden Studie sind das die drei in Kapitel 3 beschriebenen Variablen MLC, DC/C und ComNP/C.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

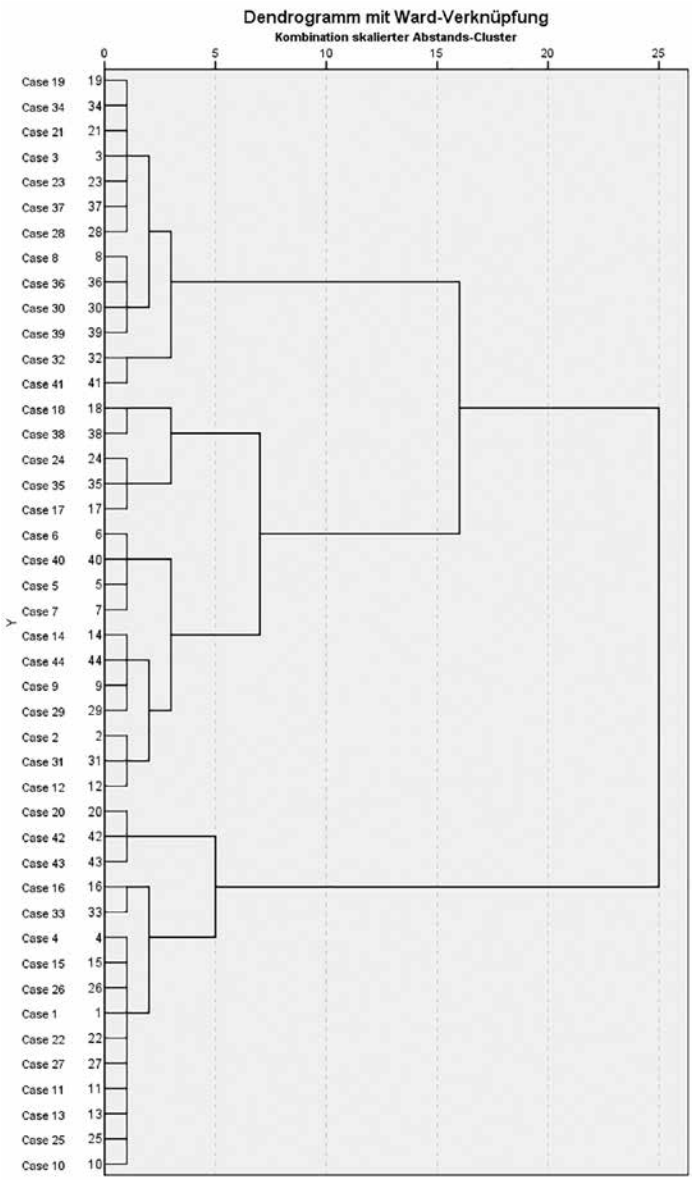


Abbildung 19: Dendrogramm der HCA zur Bestimmung der Clusterzahl in den k-means Clusteranalysen

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

für die 3-Clusterlösung bei der durchgeführten HCA. Deutlich wird, dass bei den drei Variablen ‚TP1_MLC', ‚TP1_DC/C' und ‚TP1_ComNP' deutliche Unterschiede zwischen den drei hier gebildeten Clustern bestehen. Für ‚TP1_MLC' und ‚TP1_ComNP' zeigt sich ein kontinuierlicher Anstieg über die drei Cluster, während die Variable ‚TP1_DC/C' nur in Cluster 2 einen Wert deutlich über null erreicht und in den anderen beiden Clustern gegen null tendiert.

ONEWAY ANOVA deskriptive Statistiken – 3-Clusterlösung									
		N	Mittelwert	Std.-Abweichung	Std.-Fehler	95 %-Konfidenzintervall für den Mittelwert		Minimum	Maximum
						Untergrenze	Obergrenze		
TP1_MLC	1	15	4,268	0,499	0,1296	3,991	4,545	3,330	5,000
	2	16	5,531	0,667	0,167	5,175	5,886	4,730	7,090
	3	13	5,725	0,512	0,142	5,415	6,035	5,110	6,910
	Gesamt	44	5,158	0,857	0,129	4,897	5,418	3,330	7,090
TP1_DC/C	1	15	0,008	0,032	0,008	–0,010	0,026	0,000	0,130
	2	16	0,163	0,062	0,015	0,130	0,196	0,080	0,300
	3	13	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Gesamt	44	0,062	0,088	0,013	0,036	0,089	0,000	0,300
TP1_ComNP/C	1	15	0,134	0,137	0,035	0,059	0,210	0,000	0,400
	2	16	0,274	0,161	0,040	0,188	0,359	0,000	0,580
	3	13	0,383	0,100	0,028	0,323	0,444	0,200	0,500
	Gesamt	44	0,259	0,168	0,025	0,208	0,310	0,000	0,580

Tabelle 27: Deskriptive Statistiken der einfaktoriellen ANOVA für die 3-Clusterlösung

Diese Überprüfung zeigt, dass die Teilnehmendenanzahl in Cluster 1 für alle drei Clusterlösungen (also bei der Bildung einer 2-, 3- oder 4-Clusterlösung) konstant bleibt (n=15). Die 3-Clusterlösung differenziert Cluster 2 aus der 2-Clusterlösung genauer aus, während sich in der 4-Clusterlösung ein relativ kleines viertes Cluster (n=5) bildet. Insgesamt bilden sich in der 3-Clusterlösung Cluster, die in ihrer Größe vergleichbar sind. Die Varianzanalyse zeigt darüber hinaus, dass die in den Clustern anhand der ausgewählten Clustervariablen gebildeten Gruppen sich signifikant voneinander unterscheiden (vgl. Tabelle 28).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Zusammenfassung: Einfaktorielle ANOVA – 3-Clusterlösung						
		Quadrat-summe	df	Mittel der Quadrate	F	Signi-fikanz
TP1_MLC	Zwischen den Gruppen	18,283	2	9,141	28,137	0,000
	Innerhalb der Gruppen	13,320	41	0,325		
	Gesamt	31,603	43			
TP1_DC/C	Zwischen den Gruppen	0,258	2	0,129	73,414	0,000
	Innerhalb der Gruppen	0,072	41	0,002		
	Gesamt	0,329	43			
TP1_ComNP/C	Zwischen den Gruppen	0,438	2	0,219	11,649	0,000
	Innerhalb der Gruppen	0,771	41	0,019		
	Gesamt	1,209	43			

Tabelle 28: Zusammenfassung der einfaktoriellen ANOVA für die 3-Clusterlösung

Post-Hoc-Tests (Tukey) mit zusätzlichen 1000 Bootstraps zeigen darüber hinaus, welche Gruppen in den einzelnen Clustern sich für die drei Clustervariablen signifikant unterscheiden (vgl. Tabelle 29).

Bootstrap für Mehrfachvergleiche								
	Tukey-HSD							
Abhängige Variable	(I) 3-cluster solution	(J) 3-cluster solution	Mittlere Differenz (I–J)	Sig.	Bootstrapa			
					Verzer-rung	Std.-Fehler	BCa 95 % Konfi-denzintervall	
							Unterer Wert	Oberer Wert
TP1_MLC	1	2	–1,262*	0,000	0,005	0,209	–1,740	–0,811
		3	–1,457*	0,000	–0,012	0,194	–1,826	–1,135
	2	1	1,263*	0,000	–0,005	0,209	0,891	1,654
		3	–0,194	0,635	–0,017	0,220	–0,589	0,178
	3	1	1,457*	0,000	0,012	0,194	1,099	1,880
		2	0,194	0,635	0,017	0,220	–0,281	0,697
TP1_DC/C	1	2	–0,155*	0,000	0,000	0,017	–0,188	–0,122
		3	0,008	0,860	0,000	0,008	0,000	0,025
	2	1	0,155*	0,000	–0,000	0,017	0,121	0,188
		3	0,163*	0,000	0,000	0,015	0,136	0,194
	3	1	–0,008	0,860	–0,000	0,008	–0,028	0,000
		2	–0,163*	0,000	–0,000	0,015	–0,195	–0,135

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Bootstrap für Mehrfachvergleiche								
	Tukey-HSD							
Abhängige Variable	(I) 3-cluster solution	(J) 3-cluster solution	Mittlere Differenz (I–J)	Sig.	Bootstrapa			
					Verzerrung	Std.-Fehler	BCa 95 % Konfidenzintervall	
							Unterer Wert	Oberer Wert
TP1_ComNP/C	1	2	–0,139*	0,019	0,004	0,055	–0,257	–0,019
		3	–0,249*	0,000	0,003	0,045	–0,337	–0,154
	2	1	0,139*	0,019	–0,004	0,055	0,038	0,234
		3	–0,110	0,094	–0,001	0,048	–0,204	–0,017
	3	1	0,249*	0,000	–0,003	0,045	0,155	0,334
		2	0,110	0,094	0,001	0,048	0,013	0,210
a Sofern nicht anders angegeben, beruhen die Bootstrap-Ergebnisse auf 1000 Bootstrap-Stichproben * Die Differenz der Mittelwerte ist auf dem Niveau 0.05 signifikant.								

Tabelle 29: Tukey Post-Hoc-Tests für Mehrfachvergleiche plus Bootstrap

Zusammengenommen mit den Ergebnissen der Analyse der Zuordnungsübersicht und der Informationen aus dem Dendrogramm scheint in diesem Fall eine 3-Clusterlösung geeignet zu sein, um möglichst viele Informationen der Teilnehmenden in einer longitudinalen k-means-Clusteranalyse abzubilden. Nachdem durch eine initiale HCA diese 3-Clusterlösung identifiziert wurde, wird zu jedem Zeitpunkt (TP1, TP2, TP3) eine k-means Clusteranalyse durchgeführt, wobei jeweils die Clusterzentren der Analyse zum vorangegangenen Zeitpunkt die initialen Clusterzentren für die darauffolgende Clusteranalyse bilden. Die TN können die Zughörigkeit zu einem Cluster für jeden TP wechseln. Die Ergebnisse der k-means Analyse beschreiben also die Zusammensetzung der TN in den Clustern sowie die Clusterzentren bzw. Mittelwerte der Variablen aller dem Cluster zugehörigen TN. Da die HCA als initialer Schritt durchgeführt wird, um eine geeignete Clusteranzahl für die folgenden Analysen festzulegen, werden an dieser Stelle nur die Zusammensetzungen der Cluster beschrieben, die durch die k-means Analysen zu den Zeitpunkten 1, 2 und 3 gebildet werden[208]:

208 Angegeben werden hier einerseits die Fälle pro Cluster zu jedem Zeitpunkt und andererseits die finalen Clusterzentren, die jede Clusterlösung ergeben hat.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Zusammensetzung der Cluster zu Zeitpunkt 1:

Anzahl der Fälle in jedem Cluster		
Cluster	1	11
	2	24
	3	9
Gültig		44
Fehlend		0

Tabelle 30: Anzahl der Fälle in jedem Cluster zu Zeitpunkt 1 (TP1)

Zu Zeitpunkt 1 (TP1) befinden sich elf Lernende in Cluster 1 und neun Lernende in Cluster 3. Das größte Cluster ist zu TP1 Cluster 2 mit insgesamt 24 der Teilnehmenden.

TP1 Clusterzentren			
	Cluster		
	1	2	3
TP1_MLC	4,07	5,22	6,33
TP1_DC/C	0,01	0,07	0,10
TP1_ComNP/C	0,17	0,30	0,25

Tabelle 31: Clusterzentren zu Zeitpunkt 1 (TP1). Gezeigt werden die Clusterzentren für die durchschnittliche Länge der MSE (MLC), die Subordinationsrate (DC/C) und die komplexen Nominalphrasen pro MSE (ComNP/C).

Bei der Analyse der Clusterzentren (siehe Tabelle 31) fällt auf, dass deutliche Unterschiede zwischen den Clustern bei den Variablen MLC und DC/C bestehen. Für diese beiden Variablen steigen die Werte jeweils linear in den drei Clustern. Bei der Variable ComNP/C findet sich der höchste Wert in Cluster 2. Insgesamt scheinen sich die Teilnehmenden zu Zeitpunkt 1 vor allem anhand der durchschnittlichen Länge von satzwertigen Einheiten und anhand der Subordinationsrate in distinktive Gruppen einteilen zu lassen. Bei den komplexen Nominalphrasen ist keine lineare Steigerung zwischen den Clustern zu erkennen. Anzumerken ist hier sicherlich, dass sich sowohl die Werte für Subordination als auch die Werte für komplexe Nominalphrasen auf sehr niedrigem Niveau bewegen. Lernende auf diesem Niveau produzieren nur sehr wenige Nebensätze oder Relativsätze und komplexe Nominalphrasen treten insgesamt noch selten auf. Cluster 1 zeichnet sich im Vergleich durch kürzere satzwertige Einheiten und keine untergeordneten Strukturen aus. Der Wert für die komplexen Nominalphrasen ist der niedrigste in den drei Clustern. Insgesamt scheint es sich bei dieser Gruppe zu TP1 um eher leistungsschwächere Lernende zu handeln, die noch unsicher mit vielen Strukturen arbeiten. In Cluster 2 finden sich Lernende, die etwas längere satzwertige Einheiten produzieren, aber fast keine subordinierenden Strukturen. Der Wert für die komplexen Nominalphrasen ist in diesem Cluster am höchsten, aber immer noch niedrig (0,3). Nur in jeder dritten satzwertigen Einheit findet sich eine komplexe Nominalphrase.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Zeitpunkt 1

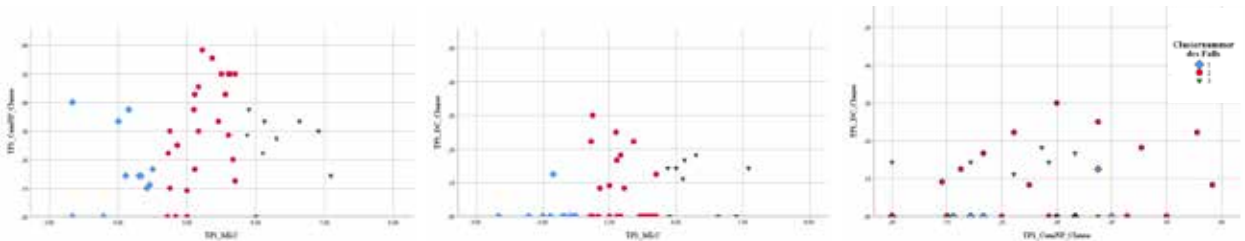


Abbildung 20: Gruppierte Streudiagramme zu Zeitpunkt 1; MLC-ComNP/C, MLC-DC/C, ComNP/C-DC/C

Zeitpunkt 2

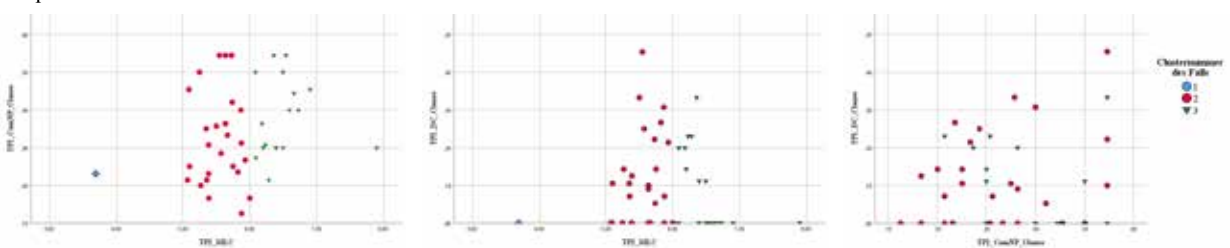


Abbildung 21: Gruppierte Streudiagramme zu Zeitpunkt 2; MLC-ComNP/C, MLC-DC/C, ComNP/C-DC/C

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Zeitpunkt 3

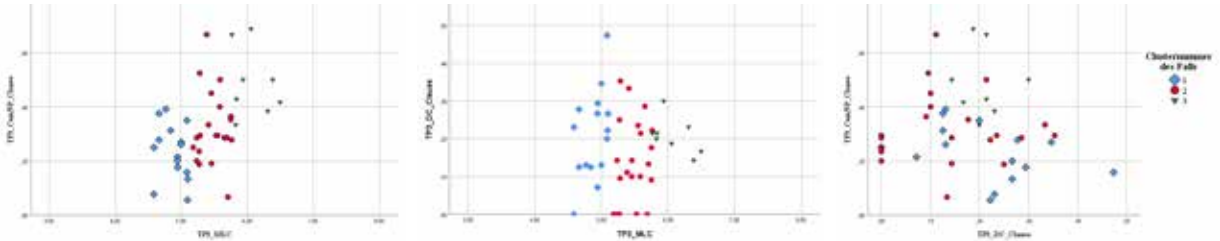


Abbildung 22: Gruppierte Streudiagramme zu Zeitpunkt 3; MLC-ComNP/C, MLC-DC/C, ComNP/C-DC/C

Zusammensetzung der Cluster zu Zeitpunkt 2:

TP2 Anzahl der Fälle in jedem Cluster		
Cluster	1	1
	2	26
	3	17
Gültig		44
Fehlend		0

Tabelle 32: Anzahl der Fälle in jedem Cluster zu Zeitpunkt 2 (TP2)

Zu Zeitpunkt 2 ändert sich die Zusammensetzung der Cluster grundlegend. Zunächst fällt auf, dass nur noch ein Teilnehmender in Cluster 1 sortiert wird, die Größe von Cluster 2 bleibt etwa konstant (n=26), während sich in Cluster 3 nun fast doppelt so viele Teilnehmende finden (n=17).

TP2 Clusterzentren			
	Cluster		
	1	2	3
TP2_MLC	3,69	5,55	6,51
TP2_DC/C	0,00	0,12	0,09
TP2_ComNP/C	0,23	0,32	0,38

Tabelle 33: Clusterzentren zu Zeitpunkt 2 (TP2). Gezeigt werden die Clusterzentren für die durchschnittliche Länge der MSE (MLC), die Subordinationsrate (DC/C) und die komplexen Nominalphrasen pro MSE (ComNP/C).

Ebenso deutliche Unterschiede zu TP1 ergeben sich bei der Analyse der Clusterzentren für die drei untersuchten Variablen. Die Person in Cluster 1 schreibt noch kürzere satzwertige Einheiten als der Mittelwert dieses Clusters zu TP1. Die Mittelwerte für die Länge satzwertiger Einheiten in Cluster 2 und 3 bleiben ähnlich groß wie zu TP1. Die Person in Cluster 1 produziert keine Nebensätze, weshalb der Wert für die Subordinationsrate bei 0 liegt. Während der Wert für die Subordinationsrate zu TP1 in Cluster 2 etwas höher ist als in Cluster 3, ist dies zu TP2 nun umgedreht. Beide liegen geringfügig über den Werten zu TP1. Insgesamt etwas höher liegen die Mittelwerte für die Rate komplexer Nominalphrasen pro MSE. Wie zu Zeitpunkt 1 scheint zu Zeitpunkt 2 vor allem die Länge satzwertiger Einheiten zwischen den Teilnehmenden zu differenzieren.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Zusammensetzung der Cluster zu Zeitpunkt 3:

TP3 Anzahl der Fälle in jedem Cluster		
Cluster	1	15,000
	2	21,000
	3	8,000
Gültig		44,000
Fehlend		0

Tabelle 34: Anzahl der Fälle in jedem Cluster zu Zeitpunkt 2 (TP2)

Die Clustergrößen zu TP3 unterscheiden sich deutlich von den Clustergrößen zu TP2. Cluster 1 beinhaltet nun 15 Teilnehmende. Cluster 2 bleibt mit 21 Teilnehmenden das größte Cluster. Cluster 3 ist im Vergleich zu TP2 etwas kleiner geworden und umfasst zu diesem Zeitpunkt acht Teilnehmende.

TP3 Clusterzentren			
	Cluster		
	1	2	3
TP3_MLC	4,89	5,48	6,08
TP3_DC/C	0,21	0,14	0,21
TP3_ComNP/C	0,23	0,32	0,49

Tabelle 35: Clusterzentren zu Zeitpunkt 3 (TP3). Gezeigt werden die Clusterzentren für die durchschnittliche Länge der MSE (MLC), die Subordinationsrate (DC/C) und die komplexen Nominalphrasen pro MSE (ComNP/C).

Die Clusterzentren zu Zeitpunkt 3 (vgl. Tabelle 35) unterscheiden sich deutlich von denen der Clusterzusammensetzung zu Zeitpunkt 2. In Cluster 1 ist der Mittelwert für die satzwertigen Einheiten wie zu TP1 und TP2 am geringsten.

Während zu den Zeitpunkten 1 und 2 vor allem die Länge der satzwertigen Einheiten die Clusterzugehörigkeit zu bestimmen scheint, unterscheiden sich zu Zeitpunkt 3 zusätzlich die Werte für komplexe Nominalphrasen pro MSE. Hier fällt vor allem der Wert von 0,49 in Cluster 3 im Vergleich zu den beiden vorangegangenen Zeitpunkten als relativ hoch auf.

Im nächsten Schritt werden aus den Zugehörigkeiten zu den Clustern longitudinale Verlaufskurven für alle Teilnehmenden bestimmt. Es ergeben sich basierend auf den Clustergruppen zu den drei Zeitpunkten 15 verschiedene Verlaufskurven, die jeweils von unterschiedlich vielen Teilnehmenden durchlaufen werden (vgl. Tabelle 36).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

		Häufigkeit	Prozent der Teilnehmenden
Gruppenzugehörigkeiten in den Clusteranalysen (abc; a = TP1; b = TP2; c = TP3)	112	1	2,3
	121	2	4,5
	122	3	6,8
	131	2	4,5
	132	3	6,8
	221	8	18,2
	222	4	9,1
	223	4	9,1
	231	1	2,3
	232	3	6,8
	233	4	9,1
	321	1	2,3
	322	4	9,1
	331	1	2,3
	332	3	6,8
	Gesamt	44	100,0

Tabelle 36: Verlauf der Clusterzugehörigkeiten über die drei Zeitpunkte

Daran anschließend werden die vorhandenen Verlaufskurven in sieben Gruppen zusammengefasst.[209] Als gruppendistinktives Merkmal wird die Anzahl an Clusterwechseln über die drei Zeitpunkte angesetzt. Dies verfolgt das Ziel, Gruppen von Lernenden zu identifizieren, die sich in der Variabilität innerhalb ihrer Entwicklungsverläufe ähneln. Die diesem Vorgehen zugrundeliegende Annahme ist, dass Lernende, die in longitudinalen Clusteranalysen häufiger als andere Lernende ihre Clusterzugehörigkeit wechseln, variablere Entwicklungsverläufe aufweisen, d.h., dass sich in den Daten der Lernenden eine größere Variabilität widerspiegelt. Lernende, die in ihren Clusterzugehörigkeiten keine Wechsel vollziehen, weisen insgesamt stabilere Entwicklungsverläufe auf. Dabei wurden folgende Unterscheidungen gesetzt (vgl. Tabelle 36):

- Teilnehmende, die die Gruppenzugehörigkeit im Verlauf nicht ändern (n=4)
- Teilnehmende, die
  a) die Gruppenzugehörigkeit einmal wechseln von Cluster 1 → Cluster 2 (n=4)
  b) die Gruppenzugehörigkeit einmal wechseln von Cluster 2 → Cluster 3 (n=8)
  c) die Gruppenzugehörigkeit einmal wechseln von 2 → 1 (n=8)
  d) die Gruppenzugehörigkeit einmal wechseln von 3 → 2 (n=7)
- Teilnehmende, die die Gruppenzugehörigkeit zweimal wechseln und dabei in zwei verschiedene Cluster eingeordnet wurden (n=8)
- Teilnehmende, die die Gruppenzugehörigkeit zweimal wechseln und dabei in drei verschiedene Cluster eingeordnet wurden (n=5)

209 Dieses Vorgehen lehnt sich methodisch an Pinel / Csizér (2015) an.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Verlaufskurve anhand der Clusteranalysen					
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Kein Wechsel der Cluster	4	9,1	9,1	9,1
	Ein Wechsel von 1 → 2 (112; 122)	4	9,1	9,1	18,2
	Ein Wechsel von 2 → 3 (223; 233)	8	18,2	18,2	36,4
	Ein Wechsel von 2 →1 (211; 221)	8	18,2	18,2	54,5
	Ein Wechsel von 3 → 2 (322; 332)	7	15,9	15,9	70,5
	Zwei Wechsel, zwei Cluster	8	18,2	18,2	88,6
	Zwei Wechsel, drei Cluster	5	11,4	11,4	100,0
	Gesamt	44	100,0	100,0	

Tabelle 37: Verlaufskurve anhand der Clusteranalysen – Gruppenbildung

Um die grundsätzliche Annahme eher variablerer und eher stabilerer Lernverläufe weiter zu untersuchen, werden nun die Daten der Clustervariablen der variableren und stabileren Gruppen aggregiert und grafisch dargestellt. Diese grafischen Repräsentationen sollen im nächsten Schritt genauer betrachtet werden. In den Abbildungen 24, 25 und 26 sind jeweils die Verlaufskurven der sieben longitudinalen Entwicklungscluster für die drei Clustervariablen anhand der jeweiligen Mittelwerte zu sehen. Zusätzlich sind in den Diagrammen 95 % Konfidenzintervalle der Standardfehler zu sehen.

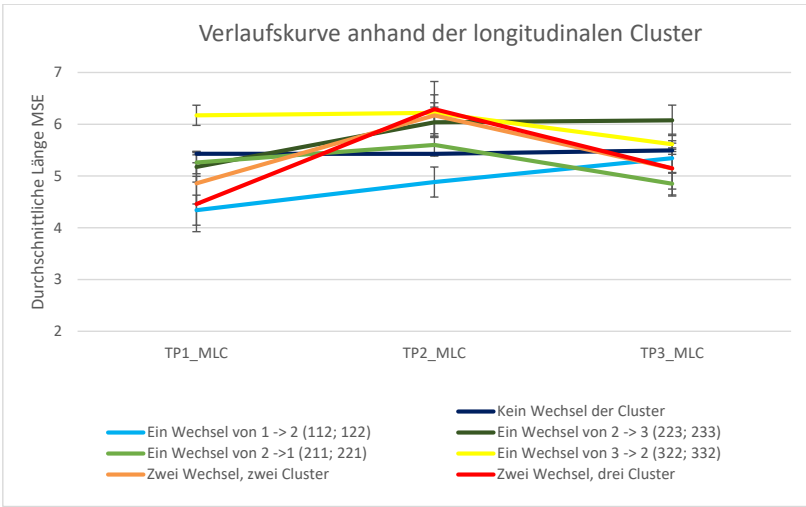


Abbildung 23: Verlaufskurve anhand der longitudinalen Cluster für die Variable ‚Durchschnittliche Länge einer MSE'; gezeigt werden jeweils die Mittelwerte der einzelnen Gruppen zu den drei Zeitpunkten.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

In Abbildung 23 sind anhand der Mittelwerte die Verlaufskurven für die sieben longitudinalen Cluster zu sehen. Auf der x-Achse wird die durchschnittliche Länge einer MSE abgebildet. Im Vergleich der sieben longitudinalen Clustergruppen hinsichtlich der durchschnittlichen Länge einer MSE wird deutlich, dass die Verlaufskurven sich relativ deutlich voneinander unterscheiden. Die Werte der Gruppe, die über die drei Zeitpunkte das Cluster gar nicht wechselt (dunkelblaue Linie), bleiben konstant bei etwa 5,5 Wörtern pro MSE. Die Werte der vier Gruppen, die einmal die Clusterzugehörigkeit wechseln, fallen oder steigen leicht, zeigen aber insgesamt eine relative Stabilität. Für die Werte der beiden Gruppen, die zweimal die Clusterzugehörigkeit ändern, fällt ein Muster auf, bei dem die Werte von Zeitpunkt 1 zu Zeitpunkt 2 stark steigen und dann zu Zeitpunkt 3 wieder deutlich abfallen. Dabei ähneln sich diese beiden Gruppen stark. Für die anderen Gruppen ist ein solches pyramidenförmiges Entwicklungsmuster nicht rekonstruierbar. Die Änderungen der Werte zwischen den Zeitpunkten sind für diese beiden Gruppen so stark, dass sich die Hypothese von stabileren und variableren Gruppen, die durch die Clusteranalysen identifiziert wurden, zu bestätigen scheint.

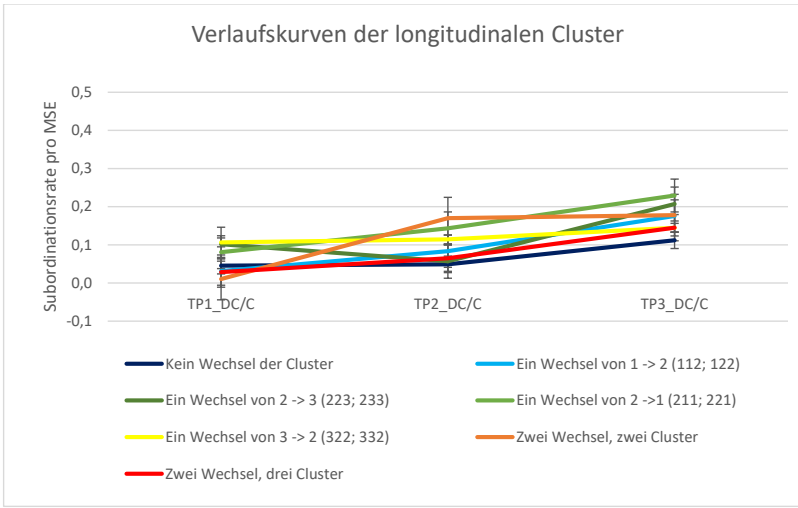


Abbildung 24: Verlaufskurve anhand der longitudinalen Cluster für die Variable ‚Subordinationsrate pro MSE'; gezeigt werden jeweils die Mittelwerte der einzelnen Gruppen zu den drei Zeitpunkten.

Für die Verlaufskurven hinsichtlich der ‚Subordinationsrate pro MSE' (vgl. Abbildung 24) lassen sich nicht so deutliche Unterschiede zwischen den einzelnen Clustergruppen herausarbeiten wie für die ‚Durchschnittliche Länge einer MSE'. Keine der Verlaufskurven zeichnet sich durch eine auffallend große Variabilität

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

aus. Die Werte aller Kurven liegen zu Zeitpunkt 3 höher als zu Zeitpunkt 1. Zusätzlich fallen einige linear ansteigende Verlaufskurven auf, wie etwa das longitudinale Cluster 5 (Ein Wechsel von 3 → 2; gelbe Linie).

Für die Werte zur Variablen ‚Rate komplexer Nominalphrasen pro MSE' stellt sich die Situation ähnlich dar (vgl. Abbildung 25). Alle Werte bewegen sich in einem vergleichbaren Fenster und für keine der Verlaufskurven lässt sich mithilfe der Visualisierung eine sehr große Variabilität herausarbeiten. Im Vergleich zu den Werten für die ‚Subordinationsrate' fällt auf, dass einige der Werte der Verlaufskurven zu Zeitpunkt 3 geringere Werte aufweisen als zu Zeitpunkt 1. Dies ist etwa für die Werte der Gruppe 1 der Fall, bei der es zu keinerlei Wechsel der Cluster über die drei Zeitpunkte kommt (dunkelblaue Linie), aber ebenso für die Werte der Gruppe, die zweimal die Cluster wechselt und dabei in zwei verschiedene Cluster eingeordnet wurde (orangene Linie). Auffällig ist, dass sechs der sieben Gruppen zu Zeitpunkt 3 fast identische Werte aufweisen. Nur die Werte der Gruppe, die einmal die Clusterzugehörigkeit wechselt, und zwar von Cluster 2 zu Cluster 3 (dunkelgrüne Linie), liegen etwas höher.

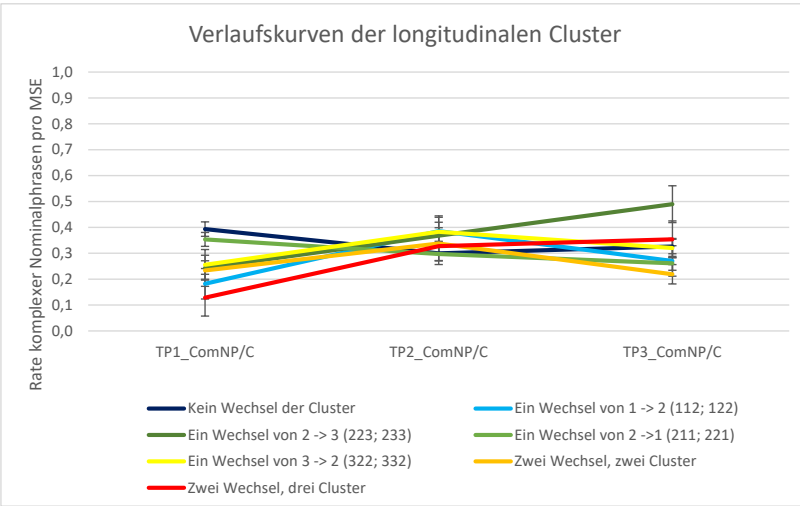


Abbildung 25: Verlaufskurve anhand der longitudinalen Cluster für die Variable ‚Rate komplexer Nominalphrasen pro MSE'; gezeigt werden jeweils die Mittelwerte der einzelnen Gruppen zu den drei Zeitpunkten.

Die hier gesammelten, ausschließlich deskriptiven Beobachtungen stützen teilweise die Hypothese, dass es durch die Clusteranalysen gelungen ist, Gruppen von Lernenden zu identifizieren, die sich in ihrer Entwicklung hinsichtlich der in den Daten vorhandenen Variabilität ähneln. Ausgesprochen deutlich wurde dies für die Werte der Variable ‚Durchschnittliche Länge einer MSE', die, wie oben

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

beschrieben, gleichzeitig die Clustervariable darstellt, die zu Zeitpunkt 1 und 2 am ehesten zwischen der Zugehörigkeit zu einem Cluster differenziert. Für die beiden Variablen ‚Subordinationsrate pro MSE' und ‚Rate komplexer Nominalphrasen' ist kein direkter Zusammenhang zwischen Gruppenzugehörigkeit und variableren Verlaufskurven anhand der Datenvisualisierungen zu rekonstruieren.

Überprüfung der Ergebnisse der Clusteranalysen hinsichtlich der experimentellen Manipulation

Um in einem letzten Schritt in Bezug auf die Clusteranalysen zu überprüfen, inwiefern die experimentelle Manipulation unter Umständen einen Einfluss auf die Ergebnisse haben könnte, werden zwei Analysen durchgeführt. Einerseits zweifache ANOVAs mit Messwiederholung (*2-way repeated measures ANOVA*) anhand der drei Clustervariablen. Andererseits werden die sieben über die Clusteranalysen ermittelten longitudinalen Cluster mithilfe eines Chi Quadrat-Tests daraufhin überprüft, ob ein Zusammenhang zwischen der Zugehörigkeit zu einer der Experimentalgruppen und der Einordnung in eines der sieben longitudinalen Cluster besteht.

Zunächst werden die Ergebnisse der ANOVAs berichtet. Für die ANOVAs werden jeweils die drei Zeitpunkte als Faktoren der Messwiederholung gewählt (*repeated measures factors*). Die Zuteilung zu einer der beiden Instruktionsgruppen (Syntax oder Semantik; vgl. Kapitel 3.1.5 für einen Überblick über die Unterschiede zwischen den beiden Gruppen) dient als Zwischensubjektfaktor (*Between-Subject-Factor*). Im Folgenden werden nur die Ergebnisse berichtet, die sich auf mögliche Effekte der ‚Instruktion' beziehen. Effekte, die sich ausschließlich auf Unterschiede zwischen den Zeitpunkten oder auf die Interaktion der Effekte zwischen ‚Zeitpunkt' und ‚Instruktion' für die gesamte Gruppe beziehen, werden an dieser Stelle nicht berichtet. Effekte werden dann als signifikant eingestuft, wenn der berechnete *p*-Wert kleiner als 0,001 ist. Zusätzlich wird das partielle η^2 als Effektstärke angegeben. Für jede der Analysen werden Post-hoc-Vergleiche mit Bonferroni-Korrektur durchgeführt. Die vollständigen Ergebnisse der ANOVAs und der Post-hoc-Tests finden sich im Anhang. In Abbildung 26 sind die Mittelwerte der drei Clustervariablen über die drei untersuchten Zeitpunkte und gruppiert nach Experimentalgruppe zu sehen. In allen drei Diagrammen zeigen sich sehr ähnliche Entwicklungsmuster, was die beiden Gruppen angeht. Auffällig ist, dass die Syntaxgruppe in allen drei Komplexitätsmaßen zu Zeitpunkt 1 etwas höhere Werte aufweist. Für die durchschnittliche Länge einer MSE zeigt sich, dass die Mittelwerte zunächst ansteigen und zu Zeitpunkt 3 in beiden Gruppen wieder abfallen. Gleichzeitig steigen die Werte der Subordinationsrate. Die Steigung der Werte für die Rate komplexer Nominalphrasen pro MSE steigt zunächst an und flacht dann für die Syntaxgruppe ab. Für die Semantikgruppe gehen die Werte zu Zeitpunkt 3 leicht zurück. Im Folgenden werden für alle drei Komplexitätsmaße die Ergebnisse der statistischen Tests und die deskriptiven Statistiken in tabellarischer Form berichtet.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

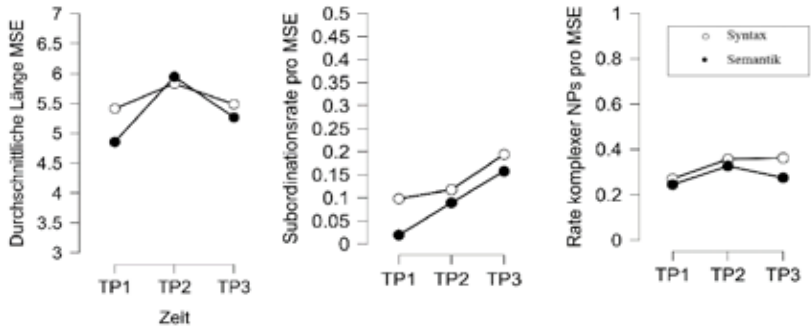


Abbildung 26: Mittelwertvergleich der drei Clustervariablen, gruppiert nach Instruktion

‚Durchschnittliche Länge einer minimalen satzwertigen Einheit'

Die deskriptive Statistik für die Variable ‚Durchschnittliche Länge einer minimalen satzwertigen Einheit' hinsichtlich möglicher Gruppenunterschiede zwischen der Syntax- und der Semantikgruppe findet sich in Tabelle 38.

Deskriptive Statistik						
	TP1_MLC		TP2_MLC		TP3_MLC	
	Syntax	Semantik	Syntax	Semantik	Syntax	Semantik
Gültig	24	20	24	20	24	20
Fehlend	0	0	0	0	0	0
Median	5,430	4,790	5,655	5,965	5,570	5,240
Mittelwert	5,411	4,854	5,828	5,943	5,486	5,264
Std,-Fehler Mittelwert	0,162	0,190	0,096	0,190	0,100	0,095
Standardabweichung	0,794	0,852	0,471	0,849	0,489	0,427
Varianz	0,630	0,726	0,222	0,721	0,239	0,182
Minimum	3,330	3,330	5,250	3,690	4,590	4,670
Maximum	7,090	6,910	6,730	7,900	6,390	6,500
25. Perzentil	4,965	4,262	5,387	5,673	5,157	4,987
50. Perzentil	5,430	4,790	5,655	5,965	5,570	5,240
75. Perzentil	5,910	5,488	6,308	6,273	5,760	5,453

Tabelle 38: Deskriptive Statistik zum Vergleich der Instruktionsgruppen: ‚Durchschnittliche Länge einer MSE'

Die Zusammenfassung der Effekte für die ANOVA ist in Tabelle 39 zu sehen. Die Post-hoc-Vergleiche mit Bonferroni-Korrektur werden in Tabelle 40 gezeigt.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Zwischensubjekteffekte						
Fälle	**Quadrat-summe**	**df**	**Mittlere Quadrat-summe**	**F**	**p**	**η^2p**
Instruktion	1,605	1	1,605	2,917	0,095	0,065
Residuals	23,117	42	0,550			
Hinweis. Typ III Quadratsumme						

Tabelle 39: Zusammenfassung der zweifachen ANOVA mit Messwiederholung: ‚Durchschnittliche Länge einer MSE'

Post-hoc Vergleich mit Bonferroni-Korrektur

		95 % KI für Mittelwertdifferenz						
		Diff	**Untere**	**Obere**	**Stan-dard-fehler**	**t**	**Cohens d**	**p_{Bonf}**
Syntax	Semantik	0,221	0,040	0,483	0,130	1,708	0,332	0,095

Tabelle 40: Post-hoc Vergleich mit Bonferroni-Korrektur zur zweifachen ANOVA mit Messwiederholung: ‚Durchschnittliche Länge einer MSE'

Damit lassen sich für die durchgeführten ANOVAs hinsichtlich der durchschnittlichen Länge einer MSE folgende Ergebnisse festhalten:

- Für die Variable ‚Durchschnittliche Länge einer minimalen satzwertigen Einheit' ergibt die Analyse keinen signifikanten Effekt in Bezug auf die Instruktion, $F(1, 42) = 2{,}917$, $p = 0{,}095$. Zusätzlich ergibt sich eine kleine Effektstärke, $\eta^2p = 0{,}065$.
- Post-Hoc-Tests mit Bonferroni-Korrektur zeigen für die ‚Durchschnittliche Länge einer MSE' keinen signifikanten Unterschied zwischen der Syntax- und der Semantikgruppe über die drei Zeitpunkte und nur eine geringe Effektstärke ($p = 0{,}095$, 95 % KI [−0,040; 0,483], $d = 0{,}332$).

‚Subordinationsrate pro MSE'

Das zweite Komplexitätsmaß, das hinsichtlich möglicher Effekte der experimentellen Manipulation analysiert wird, ist die ‚Subordinationsrate pro MSE'. In Tabelle 40 wird ein Überblick über die deskriptiven Statistiken gegeben.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Deskriptive Statistik						
	TP1_DC/Ce		TP2_DC/C		TP3_DC/C	
	Syntax	Semantik	Syntax	Semantik	Syntax	Semantik
Gültig	24	20	24	20	24	20
Fehlend	0	0	0	0	0	0
Median	0,100	0,000	0,105	0,000	0,210	0,140
Mittelwert	0,098	0,019	0,117	0,089	0,194	0,158
Std.-Fehler Mittelwert	0,018	0,014	0,025	0,025	0,024	0,022
Standardabweichung	0,090	0,062	0,123	0,114	0,116	0,098
Varianz	0,008	0,004	0,015	0,013	0,013	0,010
Minimum	0,000	0,000	0,000	0,000	0,000	0,000
Maximum	0,300	0,250	0,450	0,330	0,470	0,330
25. Perzentil	0,000	0,000	0,000	0,000	0,125	0,100
50. Perzentil	0,100	0,000	0,105	0,000	0,210	0,140
75. Perzentil	0,170	0,000	0,155	0,203	0,247	0,228

Tabelle 41: Deskriptive Statistik zum Vergleich der Instruktionsgruppen: ‚Subordinationsrate pro MSE'

Darauf folgen die Zusammenfassungen der Analyseergebnisse (vgl. Tabelle 42) und der Post-hoc-Tests (vgl. Tabelle 43).

Zwischensubjekteffekte						
Fälle	**Quadrat-summe**	**df**	**Mittlere Quadrat-summe**	**F**	**p**	**η^2p**
Instruktion	0,075	1	0,075	5,433	0,025	0,115
Residuals	0,580	42	0,014			
Hinweis. Typ III Quadratsumme						

Tabelle 42: Zusammenfassung der zweifachen ANOVA mit Messwiederholung: ‚Subordinationsrate pro MSE'

Post-hoc-Vergleich mit Bonferroni-Korrektur

95 % KI für Mittelwertdifferenz								
		Diff	**Untere**	**Obere**	**Stan-dard-fehler**	**t**	**Cohens d**	**p_{Bonf}**
Syntax	Semantik	0,048	0,006	0,089	0,021	2,331	0,463	0,025

Tabelle 43: Vergleich mit Bonferroni-Korrektur zur zweifachen ANOVA mit Messwiederholung: ‚Subordinationsrate pro MSE'

Für die ‚Subordinationsrate pro MSE' ergeben sich aus den durchgeführten Analysen folgende Ergebnisse:

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

- Für die Variable ‚Subordinationsrate pro MSE' ist kein signifikanter Effekt in Bezug auf die Instruktion zu ermitteln, $F(1, 42) = 5{,}433$, $p = 0{,}025$. Zusätzlich ergibt sich eine kleine Effektstärke, $\eta^2p = 0{,}115$.
- Post-Hoc-Tests mit Bonferroni-Korrektur zeigen für die ‚Subordinationsrate pro MSE' keinen signifikanten Unterschied zwischen der Syntax- und der Semantikgruppe über die drei Zeitpunkte und nur eine geringe Effektstärke ($p = 0{,}055$, 95 % KI [−0,040; 0,483], $d = 0{,}463$).

‚Rate komplexer Nominalphrasen pro MSE'

Die letzte der hier überprüften syntaktischen Variablen ist die ‚Rate komplexer Nominalphrasen pro MSE'. Wie zuvor findet sich in Tabelle 44 zunächst eine Zusammenfassung der relevanten deskriptiven Statistiken:

Deskriptive Statistik						
	TP1_ComNP/C		TP2_ComNP/C		TP3_ComNP/C	
	Syntax	Semantik	Syntax	Semantik	Syntax	Semantik
Gültig	24	20	24	20	24	20
Fehlend	0	0	0	0	0	0
Median	0,280	0,250	0,340	0,305	0,320	0,285
Mittelwert	0,270	0,244	0,357	0,327	0,362	0,275
Std,-Fehler Mittelwert	0,035	0,037	0,025	0,025	0,035	0,023
Standardabweichung	0,171	0,167	0,123	0,112	0,172	0,101
Varianz	0,029	0,028	0,015	0,012	0,029	0,010
Minimum	0,000	0,000	0,170	0,130	0,070	0,060
Maximum	0,580	0,500	0,550	0,550	0,690	0,450
25. Perzentil	0,140	0,107	0,265	0,247	0,265	0,198
50. Perzentil	0,280	0,250	0,340	0,305	0,320	0,285
75. Perzentil	0,355	0,380	0,440	0,410	0,500	0,335

Tabelle 44: Deskriptive Statistik zum Vergleich der Instruktionsgruppen: ‚Rate komplexer Nominalphrase pro MSE'

In Tabelle 45 und in Tabelle 46 werden die Ergebnisse der ANOVA und des durchgeführten Post-hoc-Vergleiches zur Variablen ‚Rate komplexer Nominalphrasen pro MSE' berichtet.

Zwischensubjekteffekte						
Fälle	**Quadrat-summe**	**df**	**Mittlere Quadrat-summe**	**F**	**p**	**η^2p**
Instruktion	0,075	1	0,075	3,701	0,061	0,081
Residuals	0,847	42	0,020			
Hinweis. Typ III Quadratsumme						

Tabelle 45: Zusammenfassung der zweifachen ANOVA mit Messwiederholung: ‚Rate komplexer Nominalphrase pro MSE'

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Post-hoc-Vergleich mit Bonferroni-Korrektur

		95 % KI für Mittelwertdifferenz						
		Diff	**Untere**	**Obere**	**Standardfehler**	**t**	**Cohens d**	**p_{Bonf}**
Syntax	Semantik	0,048	–0,002	0,098	0,025	1,924	0,329	0,061

Tabelle 46: Vergleich mit Bonferroni-Korrektur zur zweifachen ANOVA mit Messwiederholung: ‚Rate komplexer Nominalphrasen pro MSE'

Es ergeben sich für das Maß der ‚Rate komplexer Nominalphrasen pro MSE' damit folgende Ergebnisse:

- Das Komplexitätsmaß ‚Rate komplexer Nominalphrasen pro MSE' zeigt keinen signifikanten Effekt in Bezug auf die Instruktion, $F(1, 42) = 3{,}701$, $p = 0{,}061$. Zusätzlich ergibt sich eine kleine Effektstärke, $\eta^2p = 0{,}081$.
- Post-Hoc-Tests mit Bonferroni-Korrektur zeigen für die ‚Rate komplexer Nominalphrasen pro MSE' keinen signifikanten Unterschied zwischen der Syntax- und der Semantikgruppe über die drei Zeitpunkte und nur eine geringe Effektstärke ($p = 0{,}061$, 95 % KI [–0,040; 0,483], $d = 0{,}329$).

Obwohl die Werte der Syntaxgruppe für alle drei Komplexitätsmaße leicht über den Werten der Semantikgruppe zu liegen scheinen, ergibt sich für keine der geprüften Variablen und zu keinem Zeitpunkt ein statistisch signifikanter Effekt, der auf Gruppenunterschiede hindeuten würde.

Um zusätzlich zu überprüfen, ob die Gruppenzugehörigkeit zu einer der Experimentalgruppen mit der Einordnung in eine der sieben longitudinalen Clustergruppen zusammenhängt, wurde ein Chi Quadrat-Test durchgeführt. Die Kreuztabelle mit der Übersicht über die Verteilung findet sich in Tabelle 47. Zusätzlich wird in Abbildung 27 die Verteilung in den einzelnen Clustern mithilfe von Kreisdiagrammen visualisiert (vgl. Abbildung 27).

Kreuztabelle: Longitudinale Clusterzugehörigkeit – Experimentalgruppe				
		Instruktion		
Longitudinale Clusterzugehörigkeit		**Syntax**	**Semantik**	**Gesamt**
Kein Wechsel der Cluster	Anzahl	3,000	1,000	4,000
	Erwartete Anzahl	2,182	1,818	4,000
	% von Gesamt	6,818 %	2,273 %	9,091 %
Ein Wechsel von 1 → 2 (112; 122)	Anzahl	1,000	3,000	4,000
	Erwartete Anzahl	2,182	1,818	4,000
	% von Gesamt	2,273 %	6,818 %	9,091 %

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Kreuztabelle: Longitudinale Clusterzugehörigkeit – Experimentalgruppe				
		Instruktion		
Longitudinale Clusterzugehörigkeit		Syntax	Semantik	Gesamt
Ein Wechsel von 2 → 3 (223; 233)	Anzahl	6,000	2,000	8,000
	Erwartete Anzahl	4,364	3,636	8,000
	% von Gesamt	13,636 %	4,545 %	18,182 %
Ein Wechsel von 2 →1 (211; 221)	Anzahl	4,000	4,000	8,000
	Erwartete Anzahl	4,364	3,636	8,000
	% von Gesamt	9,091 %	9,091 %	18,182 %
Ein Wechsel von 3 → 2 (322; 332)	Anzahl	5,000	2,000	7,000
	Erwartete Anzahl	3,818	3,182	7,000
	% von Gesamt	11,364 %	4,545 %	15,909 %
Zwei Wechsel, zwei Cluster	Anzahl	3,000	5,000	8,000
	Erwartete Anzahl	4,364	3,636	8,000
	% von Gesamt	6,818 %	11,364 %	18,182 %
Zwei Wechsel, drei Cluster	Anzahl	2,000	3,000	5,000
	Erwartete Anzahl	2,727	2,273	5,000
	% von Gesamt	4,545 %	6,818 %	11,364 %
Gesamt	Anzahl	24,000	20,000	44,000
	Erwartete Anzahl	24,000	20,000	44,000
	% von Gesamt	54,545 %	45,455 %	100,000 %

Tabelle 47: Kreuztabelle zum Vergleich der Zugehörigkeit zu einem longitudinalen Cluster und der Experimentalgruppe

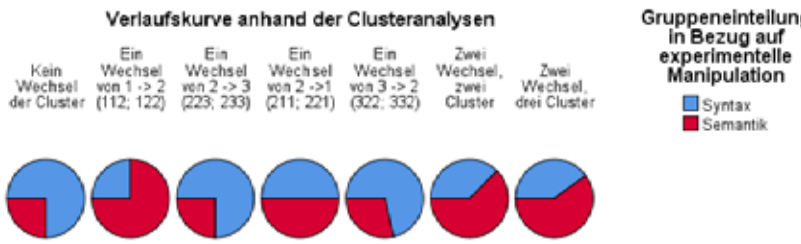


Abbildung 27: Aufteilung der longitudinalen Cluster nach Zugehörigkeit zu Experimentalgruppe

Der Chi Quadrat-Test (vgl. Tabelle 48) zeigt keinen statistisch signifikanten Zusammenhang zwischen der Instruktionsgruppe und der Zugehörigkeit zu einem der sieben longitudinalen Cluster und eine mittlere Effektstärke, $X^2(6, 44) = 5{,}669$, $p = 0{,}461$, Cramers $V = 0{,}359$.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Chi-Quadrat-Tests			
	Wert	**df**	**p**
X^2	5,669	6	0,461
N	44		
Cramers V	0,359		

Tabelle 48: Zusammenfassung des Chi Quadrat-Tests zur Überprüfung der Gruppenzugehörigkeiten

Zusammenfassend lässt sich an dieser Stelle festhalten, dass mit den hier angewendeten Verfahren und auf Grundlage der untersuchten Komplexitätsmaße keine statistisch signifikanten Unterschiede festgestellt werden, die auf die Zugehörigkeit zu einer der beiden Experimentalgruppen zurückgeführt werden kann. Dies bedeutet, dass aufgrund der hier untersuchten Maße zunächst nicht von einer Verzerrung der Ergebnisse durch die experimentelle Manipulation ausgegangen werden muss.

4.1.4 Zusammenfassung der Ergebnisse auf Gruppenebene

Auf einer allgemeinen Ebene verdeutlichten die in diesem Abschnitt herausgearbeiteten Ergebnisse, dass die auf dieselben Daten angewendeten methodischen Verfahren durchaus unterschiedliche und zum Teil scheinbar widersprüchliche Ergebnisse zeigen. Die Daten wurden einerseits auf Gruppenebene anhand der deskriptiven Statistiken und mithilfe eines Friedman-Tests analysiert. Dieser wurde eingesetzt, um zu überprüfen, ob sich die durch die Profilanalysen ermittelten Niveaustufen der Teilnehmenden insgesamt über die drei Zeitpunkte hinweg signifikant ändern bzw. unterscheiden. Zunächst schienen sowohl die deskriptiven Daten als auch die Ergebnisse des Friedman-Tests auf insgesamt eher lineare Erwerbsverläufe hinzudeuten, die zusätzlich den Annahmen implikationeller und sequenzieller Erwerbssequenzen entsprechen.

Andererseits wurden mithilfe der individuellen Lernerdaten Emergenz- und Profilanalysen durchgeführt, um herauszuarbeiten, ob die syntaktische Entwicklung der Lernenden sequenziell und implikationell anhand der vorausgesagten Erwerbsstufen verläuft. In diesen Analysen wurden die minimalen satzwertigen Einheiten und die Wortstellungsmuster, die die unterschiedlichen Profilstufen ausmachen, operationalisiert. Während die aggregierten Gruppendaten sehr deutliche Tendenzen zeigten und auf eine implikationell-sequenzielle Erwerbsfolge für die meisten der Lernenden schließen lassen, differenzierten die Emergenzanalysen auf der Ebene individueller Lernender diese Ergebnisse sehr deutlich aus. Tatsächlich scheinen diese verschiedene Facetten der Daten genauer zu perspektivieren und sich nicht zwangsläufig gegenseitig auszuschließen. Der genauere Blick auf einzelne Lernende mithilfe der Emergenzanalysen ließ die beobachteten Entwicklungen jedoch deutlich komplexer

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

erscheinen. Für viele der Lernenden konnten sich anhand der hier untersuchten Daten und aufgrund der Emergenz einzelner Strukturen zu einzelnen Zeitpunkten die für den gesteuerten und ungesteuerten Erwerb vorausgesagten Sequenzen nicht eindeutig nachweisen lassen. Die Sequenzen wurden deutlicher, wenn mehr Zeitpunkte gleichzeitig einbezogen wurden und so eine breitere Datenbasis zugrunde gelegt wurde. Gleichzeitig erlaubten die Daten es aufgrund ihrer großen Heterogenität nicht, für die hier untersuchten Lernenden (erwachsene Deutschlernende mit Arabisch als L1) alternative Erwerbsreihenfolgen der untersuchten syntaktischen Strukturen herauszuarbeiten, die überindividuell Geltung hätten. Vielmehr waren die unterschiedlichen Erwerbsverläufe durch eine sehr hohe inter- und intraindividuelle Variabilität gekennzeichnet.

Zusätzlich wurden unter Einsatz von drei ausgewählten syntaktischen Komplexitätsmaßen longitudinale Clusteranalysen über drei Zeitpunkte durchgeführt, um Gruppen von Lernerinnen und Lernern zu identifizieren, die sich in ihrer Entwicklung ähneln. Während die Lernenden zu jedem Zeitpunkt anhand der Ähnlichkeit in Bezug auf die Komplexitätsmaße gruppiert wurden, war für die Zuordnung zu longitudinalen Clustern die Variabilität in Bezug auf die Clusterzugehörigkeit das ausschlaggebende Kriterium. Ob sich durch dieses Vorgehen tatsächlich variablere und stabilere longitudinale Cluster ergaben, wurde mithilfe von Visualisierungen überprüft. Die longitudinalen Clusteranalysen zeigten einerseits, dass sich zu jedem Untersuchungszeitpunkt anhand der untersuchten Variablen ‚Durchschnittliche Länge MSE', ‚Subordinationsrate pro MSE' und ‚Rate komplexer Nominalphrasen pro MSE' distinkte Cluster bilden ließen. Vor allem das Längenmaß ‚MLC' trennte zu allen drei Zeitpunkten die Lernenden sehr deutlich in verschiedene Cluster. Darüber hinaus verdeutlichten die Clusteranalysen, dass die drei untersuchten Variablen zu den drei Zeitpunkten unterschiedlich stark zwischen den Gruppen differenzieren konnten. Während zu den ersten beiden Zeitpunkten vor allem die durchschnittliche Länge satzwertiger Einheiten zwischen den Teilnehmenden differenzierte, war es zu Zeitpunkt 3 zusätzlich die Rate komplexer Nominalphrasen pro MSE, die die einzelnen Cluster unterschied. Die Subordinationsrate schien zu keinem der Zeitpunkte in diesem eher frühen Stadium zweitsprachlicher Entwicklung für die herausgearbeiteten Cluster eine entscheidende Rolle zu spielen. Deutlich wurde, dass die Entwicklung der untersuchten Variablen alles andere als linear verläuft. Es kam beispielsweise zu Regressphänomenen. So nahm etwa die durchschnittliche Länge minimaler satzwertiger Einheiten (MLC) von Zeitpunkt 2 zu Zeitpunkt 3 sowohl auf der Ebene der gesamten Stichprobe als auch in den drei Clustern ab und ähnelte den Werten zu Zeitpunkt 1. Gleichzeitig kam es zu Steigerungen in den beiden Variablen ‚DC/C' und ‚ComNP/C', was darauf hindeuten könnte, dass komplexere Strukturen zunächst in kürzeren Einheiten produziert werden oder dass komplexere Strukturen zu kürzeren minimalen satzwertigen Einheiten führen.[210]

210 In einem nächsten Entwicklungsschritt könnte es dann zu einer Integration der komplexeren Strukturen in längere Einheiten kommen. Hier wäre es von entschei-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Darüber hinaus konnten mithilfe der longitudinalen Clusteranalyse Gruppen von Lernenden herausgearbeitet werden, die sich hinsichtlich der Variabilität innerhalb der Entwicklungsverläufe gemessen an den hier untersuchten Variablen ähneln bzw. unterscheiden. Dabei ließen sich sieben unterschiedliche longitudinale Profile rekonstruieren, die sich zwischen den Polen ‚sehr stabil' und ‚sehr variabel' einordnen lassen.

Zur Überprüfung eines möglichen Einflusses der experimentellen Manipulation wurden zweifache ANOVAs mit Messwiederholungen durchgeführt. Es ließen sich anhand der ausgewählten Stichprobe und der hier untersuchten Komplexitätsmaße keine signifikanten Unterschiede zwischen den beiden Experimentalgruppen feststellen. Zusätzlich zeigte ein Chi Quadrat-Test, dass kein statistisch signifikanter Zusammenhang zwischen der Zugehörigkeit zu einer der sieben longitudinalen Clustergruppen und den Instruktionsgruppen nachweisbar ist.

Die hier kombinierten methodischen Schritte zeigten auf, wie ein detaillierterer Blick auf die Daten weitere Bedeutungsebenen freigeben konnte und verdeutlichten zudem die Notwendigkeit bei der Analyse von Lernertexten verschiedene Parameter und unterschiedlich granulare Perspektiven anzulegen.

4.2 Ergebnisse: Einzelfallanalysen

In den folgenden Abschnitten werden die Ergebnisse der Einzelfallanalysen für die beiden Teilnehmenden TN024 und TN025 dargestellt. Zunächst werden deskriptive Statistiken berichtet. Daraufhin folgen die Ergebnisse verschiedener grafischer Analysen zur Variabilität innerhalb der Daten, zur Interaktion der verschiedenen von den Teilnehmenden produzierten Wortstellungsmuster und zu Interaktionen zwischen den syntaktischen Variablen. Diese Analysen werden durch gleitende Korrelationen (*moving correlations*) ergänzt, um die Beziehungen zwischen den Variablen zu untersuchen. Zuletzt werden in diesem Kapitel die Ergebnisse der *Change Point*-Analysen besprochen, mithilfe derer (mögliche) Entwicklungssprünge (bzw. systematische Veränderungen) in der zweitsprachlichen Entwicklung von TN024 und TN025 identifiziert werden sollen. In jedem Schritt werden die beiden Lernenden zunächst in ihrer Entwicklung gesondert betrachtet und dann die Ergebnisse der beiden Lernenden vergleichend aufeinander bezogen. Die experimentelle Manipulation (vgl. Kapitel 3.1.5) spielt in diesem Teil der vorliegenden Studie keine Rolle, da beide Lernenden über die gesamte Dauer der Studie in derselben Kursgruppe gelernt haben und sich daher keine Unterschiede, die auf der Manipulation beruhen könnten, herausarbeiten lassen. Eine Diskussion der Ergebnisse im Kontext des

dender Bedeutung, die Entwicklung der untersuchten Lernenden zu weiteren Zeitpunkten miteinzubeziehen, was im Rahmen dieser Studie leider nicht möglich war. Dies könnte helfen, diese Tendenz besser einzuschätzen und zu verfolgen.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

aufgezeigten Forschungsstandes und der Daten auf Gruppenebene erfolgt im Anschluss in Kapitel 5.

4.2.1 Deskriptive Beschreibung

Zunächst werden an dieser Stelle deskriptive Daten präsentiert, die sich aus den von TN024 und TN025 produzierten Texten im Laufe der Studie ergeben. Die deskriptiven Daten werden einmal für die gesamte Studiendauer und dann getrennt für den ersten und zweiten Studienteil angegeben (vgl. für den zeitlichen Aufbau der Studie Kapitel 3.1.3). Dabei werden deskriptive Daten für alle in dieser Studie untersuchten Variablen gezeigt (für eine Beschreibung der beiden Lernenden vgl. das Kapitel 3.3.1). Neben den deskriptiven Daten, die sich auf die zentralen Lageparameter für die ausgewählten Variablen beziehen, finden sich in Tabelle 58 und Tabelle 59 am Ende dieses Unterkapitels die Rohdaten der beiden Teilnehmenden für die einzelnen Texte.

In die vorliegenden Analysen gingen 45 Texte der beiden Lernenden (n=22 von TN024; n=23 von TN025) ein. In diesen 45 Texten wurden insgesamt 4423 Wörter und 727 minimale satzwertige Einheiten (MSE) produziert (vgl. Tabelle 49). Dabei wurden pro Text im Schnitt 98,29 Wörter geschrieben (*Mdn* = 101, SD = 42,654) und 16,16 minimale satzwertige Einheiten (MSE) (*Mdn* = 15, SD = 7,176) verwendet. Die MSE haben eine durchschnittliche Länge von 6,140 Wörtern (*Mdn* = 6,105, SD = 1,135). Die Subordinationsrate liegt im Schnitt bei 0,132 subordinierten Nebensätzen pro Hauptsatz (*Mdn* = 0,107, SD = 0,134) und der Mittelwert der Rate komplexer Nominalphrasen pro MSE liegt bei 0,347 (*Mdn* = 0,334, SD = 0,217). Die sehr hohen Standardabweichungen für die produzierten Wörter und die MSE sind isoliert betrachtet wenig aussagekräftig. Sie könnten hier ein Hinweis darauf sein, dass sich die Texte im Laufe der Studie hinsichtlich der Werte dieser Variablen stark unterscheiden und dass am Anfang kürzere und am Ende längere Texte mit deutlich mehr Wörtern produziert werden.

Deskriptive Statistik – Gesamtstudie (TN024+TN025)[211]							
		Erreichte Profilstufe	**Anzahl MSE**	**Anzahl Wörter**	**Durchschnittliche Länge MSE (MLC)**	**Subordinationsrate pro MSE**	**Rate komplexer NP pro MSE**
N	Gültig	45	45	45	45	45	45
	Fehlend	0	0	0	0	0	0
Summe			727	4423			
Mittelwert		3,09	16,16	98,29	6,140	0,132	0,347

211 Die folgenden deskriptiven Daten und die dazugehörigen Tabellen wurden mit IBM SPSS Statistics 26 berechnet und erstellt. Die Werte sind, sofern möglich, auf die dritte Stelle nach dem Komma gerundet.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Deskriptive Statistik – Gesamtstudie (TN024+TN025)[211]						
	Erreichte Profilstufe	Anzahl MSE	Anzahl Wörter	Durchschnittliche Länge MSE (MLC)	Subordinationsrate pro MSE	Rate komplexer NP pro MSE
Standardfehler des Mittelwerts	0,145	1,070	6,358	0,169	0,020	0,032
Median	3,00	15,00	101,00	6,105	0,107	0,334
Modus	3a	19	62a	5,438a	0,000	0,400
Std.-Abweichung	0,973	7,176	42,654	1,135	0,134	0,217
Varianz	0,946	51,498	1819,346	1,289	0,018	0,047
Minimum	1	7	36	3,588	0,000	0,00
Maximum	4	38	215	9,334	0,406	0,938
a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.						

Tabelle 49: Deskriptive Daten für TN024 und TN25 (aggregiert)

Tabelle 50 und Tabelle 51 zeigen die deskriptiven Daten der beiden Teilnehmenden getrennt nach Studienphase, wobei die erste Studienphase vom Beginn der Studie (23.5.2016) bis zur Prüfung des Goethe-Zertifikat B1 (12.12.2016) dauert und die zweite Studienphase mit dem ersten Messpunkt nach der B1-Prüfung (4.1.2017) beginnt und sich bis zum Ende der Studie erstreckt (28.7.2017) (zum zeitlichen Ablauf der Studie vgl. Kapitel 3.1.3).[212] An dieser Stelle wird auf eine genaue Beschreibung jeder einzelnen Tabelle verzichtet, vielmehr werden die beiden Tabellen vergleichend beschrieben.

Deskriptive Statistik Erste Studienphase (TN024+TN025)							
		Erreichte Profilstufe	Anzahl MSE	Anzahl Wörter	Durchschnittliche Länge MSE (MLC)	Subordinationsrate pro MSE	Rate komplexer NP pro MSE
N	Gültig	24	24	24	24	24	24
	Fehlend	0	0	0	0	0	0
Summe			304	1671			
Mittelwert		2,670	12,670	69,630	5,678	0,062	0,267
Standardfehler des Mittelwerts		0,187	1,170	4,812	0,220	0,021	0,038

212 Die Daten aus der ersten Studienphase beziehen sich auf die Kursgruppe 3, die TN024 und TN025 gemeinsam besuchen. In der zweiten Studienphase besuchen TN024 und TN025 ebenfalls dieselbe Kursgruppe.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Deskriptive Statistik Erste Studienphase (TN024+TN025)						
	Erreichte Profilstufe	Anzahl MSE	Anzahl Wörter	Durchschnittliche Länge MSE (MLC)	Subordinationsrate pro MSE	Rate komplexer NP pro MSE
Median	3,000	11,000	65,000	5,441	0,000	0,250
Modus	3	8a	87a	3,588	0,000	0,250
Std.-Abweichung	0,917	5,731	23,574	1,080	0,105	0,187
Varianz	0,841	32,841	555,723	1,166	0,011	0,035
Minimum	1	7	36	3,588	0,000	0,000
Maximum	4	34	122	8,700	0,300	0,688
a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt						

Tabelle 50: Deskriptive Daten für TN024 und TN25 – erste Studienphase bis B1 (aggregiert)

Deskriptive Statistik Zweite Studienphase (TN024+TN025)							
		Erreichte Profilstufe	Anzahl MSE	Anzahl Wörter	Durchschnittliche Länge MSE (MLC)	Subordinationsrate pro MSE	Rate komplexer NP pro MSE
N	Gültig	21	21	21	21	21	21
	Fehlend	0	0	0	0	0	0
Summe			423	2752			
Mittelwert		3,570	20,140	131,050	6,667	0,212	0,439
Standardfehler des Mittelwerts		0,177	1,451	7,717	0,212	0,026	0,047
Median		4,000	19,000	122,000	6,607	0,190	0,444
Modus		4	19	103a	6,438	0,316	0,474
Std.-Abweichung		0,811	6,650	35,366	0,974	0,120	0,216
Varianz		0,657	44,229	1250,748	0,948	0,0150	0,047
Minimum		1	9	62	4,760	0,0536	0,063
Maximum		4	38	215	9,333	0,406	0,938
a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.							

Tabelle 51: Deskriptive Daten für TN024 und TN25 – zweite Studienphase bis B2/C1 (aggregiert)

Der Vergleich der beiden Studienphasen für TN024 und TN025 zeigt deutliche Tendenzen. In der ersten Studienphase liegt der Mittelwert der erreichten Profilstufe bei 2,67 (*Mdn* = 3, SD = 0,917). In der zweiten Studienphase liegt der durchschnittliche Wert für die erreichte Profilstufe bei 3,57 (*Mdn* = 4, SD = 0,811). Ins-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

gesamt werden höhere Profilstufen erreicht, während die Streuung der Daten abnimmt. Dies zeigen die kleinere Standardabweichung und die abnehmende Varianz zwischen den beiden Studienphasen (Varianz in Bezug auf die erreichten Profilstufen in Phase 1: 0,841 und in Phase 2: 0,657). Es werden in der zweiten Studienphase (Mittelwert ‚Anzahl MSE' = 20,14, *Mdn* = 19, SD = 6,65) im Vergleich zur ersten Studienphase (Mittelwert ‚Anzahl MSE'= 12,67, *Mdn* = 11, SD = 5,731) von beiden Teilnehmenden mehr MSE im Schnitt produziert. Ebenso verdoppelt sich die Zahl der produzierten Wörter pro Text fast (Phase 1: Mittelwert ‚Anzahl Wörter' = 69,63; *Mdn* = 65, SD = 23,574; Phase 2: Mittelwert ‚Anzahl Wörter' = 131,05, *Mdn* = 122, SD = 35,366) und die Raten, die die Subordination und die Verwendung komplexer Nominalphrasen verdeutlichen, zeigen eine starke Erhöhung. Am geringsten fällt der Unterschied in der Länge der minimalen satzwertigen Einheiten zwischen Studienphase 1 und 2 aus (Phase 1: Mittelwert ‚MLC' = 5,678, *Mdn* = 5,441, SD = 1,08; Phase 2: Mittelwert ‚MLC' = 6,667, *Mdn* = 6,607, SD = 0,974). Gleichzeitig deuten die Werte in Bezug auf die erreichte Profilstufe daraufhin, dass die Texte komplexere Satzstrukturen bzw. Wortstellungsmuster beinhalten.

Um die vorgestellten Daten im Weiteren für jeden der beiden Lernenden nachvollziehen zu können, werden die Daten für die Lernenden TN024 und TN025 getrennt berichtet. In den Tabellen 52 und 53 werden zunächst die deskriptiven Werte für die gesamte Studie gezeigt.

Deskriptive Statistik Gesamtstudie: TN024							
		Erreichte Profilstufe	**Anzahl MSE**	**Anzahl Wörter**	**Durchschnittliche Länge MSE (MLC)**	**Subordinationsrate pro MSE**	**Rate komplexer NP pro MSE**
N	Gültig	22	22	22	22	22	22
	Fehlend	0	0	0	0	0	0
Summe			310	2000			
Mittelwert		3,00	14,09	90,91	6,453	0,100	0,4415
Standardfehler des Mittelwerts		0,208	1,032	7,487	0,255	0,0237	0,0499
Median		3,00	13,00	87,00	6,550	0,077	0,462
Modus		3	9a	62a	4,750a	0,00	0,462a
Std.-Abweichung		0,976	4,839	35,117	1,195	0,111	0,234
Varianz		0,952	23,420	1233,229	1,429	0,012	0,055
Minimum		1	7	44	4,750	0,00	0,083
Maximum		4	28	185	9,334	0,316	0,938
a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.							

Tabelle 52: Deskriptive Daten für TN024 – Gesamtstudie

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Deskriptive Statistik Gesamtstudie: TN025							
		Erreichte Profilstufe	Anzahl MSE	Anzahl Wörter	Durchschnittliche Länge MSE (MLC)	Subordinationsrate pro MSE	Rate komplexer NP pro MSE
N	Gültig	23	23	23	23	23	23
	Fehlend	0	0	0	0	0	0
Summe			417	2423			
Mittelwert		3,17	18,13	105,35	5,840	0,163	0,258
Standardfehler des Mittelwerts		0,205	1,773	10,119	0,211	0,031	0,033
Median		3,00	16,00	107,00	5,727	0,190	0,289
Modus		4	8	103a	6,438	0,000	0,125a
Std.-Abweichung		0,984	8,503	48,528	1,011	0,149	0,156
Varianz		0,968	72,300	2354,964	1,022	0,022	0,024
Minimum		1	8	36	3,588	0,00	0,00
Maximum		4	38	215	7,500	0,406	0,545

a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.

Tabelle 53: Deskriptive Daten für TN025 – Gesamtstudie

TN024 produziert in der gesamten Studie in 22 Texten insgesamt 310 MSE mit 2000 Wörtern (vgl. Tabelle 52). Für TN025 liegen die Werte in 23 Texten mit insgesamt 417 MSE und 2423 Wörtern etwas höher (vgl. Tabelle 53). Beim Vergleich der deskriptiven Daten der beiden Teilnehmenden fällt weiterhin auf, dass TN025, bezogen auf die Gesamtstudie, mehr MSE (TN024 Mittelwert ‚Anzahl MSE' = 14,09, *Mdn* = 13, SD = 0,976; TN025 Mittelwert ‚Anzahl MSE' = 18,13, *Mdn* = 16, SD 0,968) produziert und zusätzlich etwas längere Texte als TN024 schreibt (TN024 Mittelwert ‚Anzahl Wörter' = 90,91, *Mdn* = 87, SD = 25,117; TN025 Mittelwert ‚Anzahl Wörter' = 105,35, *Mdn* = 107, SD = 48,528). Wie angedeutet können die sehr großen Standardabweichungen hinsichtlich der Daten, die sich auf den gesamten Zeitverlauf beziehen, als Hinweis darauf gelesen werden, dass sich die Textlänge der geschriebenen Texte im Studienverlauf deutlich verändert. Die produzierten MSE von TN025 sind etwas kürzer als die von TN024 (TN024 Mittelwert ‚MLC' = 6,453, *Mdn* = 6,550, SD = 1,195; TN025 Mittelwert ‚MLC' = 5,840, *Mdn* = 5,727, SD = 1,011). Während die Werte in Bezug auf die Subordinationsrate ähnlich niedrig sind, fällt auf, dass TN024 etwas mehr komplexe Nominalphrasen pro MSE produziert (TN024 Mittelwert ‚Subordinationsrate' = 0,1, *Mdn* = 0,077, SD = 0,111; TN025 Mittelwert ‚Subordinationsrate' = 0,163, *Mdn* = 0,190, SD = 0,149). Die beiden Tabellen 52 und 53 zeigen erste sehr grobe Tendenzen auf, die im Folgenden für beide Teilnehmenden genauer in den Blick genommen werden.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

In den nächsten beiden Tabellen werden für TN024 die Daten der ersten Studienphase (vgl. Tabelle 54) und der zweiten Studienphase (vgl. Tabelle 55) dargestellt.

Deskriptive Statistik Studienphase 1: TN024							
		Erreichte Profilstufe	**Anzahl MSE**	**Anzahl Wörter**	**Durchschnittliche Länge MSE (MLC)**	**Subordinationsrate pro MSE**	**Rate komplexer NP pro MSE**
N	Gültig	12	12	12	12	12	12
	Fehlend	0	0	0	0	0	0
Summe			141	811			
Mittelwert		2,667	11,750	67,583	5,886	0,046	0,319
Standardfehler des Mittelwerts		0,256	0,946	4,518	0,320	0,027	0,057
Median		3,000	11,500	65,000	5,692	0,000	0,298
Modus		3,00	9a	87,00	4,75a	0,00	0,46
Std.-Abweichung		0,888	3,279	15,652	1,109	0,093	0,199
Varianz		0,788	10,750	244,992	1,229	0,009	0,039
Minimum		1	7	44	4,75	0,00	0,08
Maximum		4	19	92	8,70	0,30	0,69
a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.							

Tabelle 54: Deskriptive Daten für TN024 – erste Studienphase

Für TN024 fällt auf, dass alle Mittelwerte und Mediane in der ersten Studienphase (vgl. Tabelle 54) deutlich niedriger liegen als in der zweiten Studienphase (vgl. Tabelle 55). Es werden höhere Profilstufen in Studienphase 2 (Mittelwert ‚Erreichte Profilstufe' = 3,4, *Mdn* = 4, SD = 0,966) erreicht als in Studienphase 1 (‚Mittelwert Profilstufe' = 2,667, *Mdn* = 3, SD = 0,888). TN024 produziert in den Texten der zweiten Studienphase mehr Wörter (Mittelwert ‚Anzahl Wörter' in Studienphase 1 = 67,583, *Mdn* = 65, SD = 15,652; Mittelwert ‚Anzahl Wörter' in Studienphase 2 = 118,900, *Mdn* = 119, SD = 31,338) und mehr (Mittelwert ‚Anzahl MSE' in Studienphase 1 = 11,750, *Mdn* = 11,5, SD = 3,279; Mittelwert ‚Anzahl MSE' in Studienphase 2 = 16,9, *Mdn* = 17, SD = 5,043) und längere minimale satzwertige Einheiten (Mittelwert ‚MLC' in Studienphase 1 = 5,886, *Mdn* = 5,886, Mdn = 5,692, SD = 1,109; Mittelwert ‚MLC' in Studienphase 2 = 7,134, *Mdn* = 6,833, SD = 0,942). Im Schnitt werden in Studienphase 2 in einem größeren Anteil der satzwertigen Einheiten Strukturen mit Verbendstellung realisiert (Mittelwert ‚Subordinationsrate' in Studienphase 1 = 0,046, *Mdn* = 0,000, SD = 0,093; Mittelwert ‚Subordinationsrate' in Studienphase 2 = 0,166, *Mdn* = 0,139, SD = 0,098) und in mehr MSE finden sich komplexe Nominalphrasen (vgl. Tabellen

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

54 und 55). Ein Median von 0 bedeutet in diesem Fall, dass mehr als die Hälfte aller Werte aus dieser Datenreihe bei 0 liegen. Anders ausgedrückt hat TN024 in mehr als der Hälfte der Texte keine Strukturen mit Verbendstellung realisiert.

Deskriptive Statistik Studienphase 2: TN024							
		Erreichte Profilstufe	**Anzahl MSE**	**Anzahl Wörter**	**Durchschnittliche Länge MSE (MLC)**	**Subordinationsrate pro MSE**	Rate komplexer NP pro MSE
N	Gültig	10	10	10	10	10	10
	Fehlend	0	0	0	0	0	0
Summe			169	1189			
Mittelwert		3,400	16,900	118,900	7,134	0,166	0,586
Standardfehler des Mittelwerts		0,306	1,595	9,910	0,298	0,031	,0061
Median		4,000	17,000	119,000	6,833	0,139	0,546
Modus		4,00	17,00a	62,00a	6,11a	0,32	0,37a
Std.-Abweichung		0,966	5,043	31,338	0,942	0,098	0,191
Varianz		0,933	25,433	982,100	0,888	0,010	0,037
Minimum		1	9	62	6,11	0,06	0,37
Maximum		4	28	185	9,33	0,32	0,94
a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.							

Tabelle 55: Deskriptive Daten zu TN024 – zweite Studienphase

Ein ähnliches Bild zeigt sich für TN025, dessen deskriptive Daten aus der ersten und aus der zweiten Studienphase im Folgenden in den Tabellen 56 und 57 gezeigt werden:

Deskriptive Statistik Studienphase 1: TN025							
		Erreichte Profilstufe	**Anzahl MSE**	**Anzahl Wörter**	**Durchschnittliche Länge MSE (MLC)**	**Subordinationsrate pro MSE**	**Rate komplexer NP pro MSE**
N	Gültig	12	12	12	12	12	12
	Fehlend	0	0	0	0	0	0
Summe			163	860			
Mittelwert		2,667	13,583	71,667	5,471	0,079	0,214
Standardfehler des Mittelwerts		0,284	2,162	8,698	0,305	0,034	0,048
Median		3,000	11,000	65,000	5,324	0,000	0,178

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Deskriptive Statistik Studienphase 1: TN025						
	Erreichte Profilstufe	**Anzahl MSE**	**Anzahl Wörter**	**Durchschnittliche Länge MSE (MLC)**	**Subordinationsrate pro MSE**	**Rate komplexer NP pro MSE**
Modus	3,00	8,00	107,00	3,59a	0,00	0,00a
Std.-Abweichung	0,985	7,489	30,131	1,055	0,117	0,166
Varianz	0,970	56,083	907,879	1,114	0,014	0,027
Minimum	1	8	36	3,59	0,00	0,00
Maximum	4	34	122	7,13	0,27	0,55
a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.						

Tabelle 56: Deskriptive Daten zu TN025 – erste Studienphase

Wie für TN024 liegen für TN025 die Mittelwerte und Mediane für alle untersuchten Variablen in der ersten Studienphase (vgl. Tabelle 56) niedriger als in der zweiten Studienphase (vgl. Tabelle 57). Der Mittelwert für die Variable ‚Erreichte Profilstufe' liegt in der ersten Studienphase bei 2,667 (*Mdn* = 3, SD = 0,985) und in der zweiten Studienphase bei 3,727 (*Mdn* = 4, SD = 0,647). Während die Texte von TN025 in der ersten Phase aus durchschnittlich 71,667 Wörtern (*Mdn* = 65, SD = 30,131) und 13,583 MSE (*Mdn* = 11, SD = 7,489) aufgebaut sind, sind es in der zweiten Studienphase durchschnittlich 142,091 Wörter (*Mdn* = 130, SD = 36,550) und 23,091 MSE (*Mdn* = 21, SD = 6,745). Die satzwertigen Einheiten werden von Studienphase 1 zu Studienphase 2 etwas länger (Mittelwert ‚MLC' in Studienphase 1 = 5,471, *Mdn* = 5,324, SD = 1,055; Mittwelwert ‚MLC' in Studienphase 2 = 6,243, *Mdn* = 6,438, SD = 0,825). Die Subordinationsrate steigt von durchschnittlich 0,079 (*Mdn* = 0, SD = 0,117) in Studienphase 1 auf 0,255 (*Mdn* = 0,238, SD = 0,128) satzwertige Einheiten mit Verbendstellung pro MSE in Studienphase 2. Ebenso steigt der Wert für komplexe Nominalphrasen pro MSE von durchschnittlich 0,214 (*Mdn* = 0,178, SD = 0,166) NPs pro MSE in der ersten Phase auf 0,305 (*Mdn* = 0,316, SD = 0,137) in der zweiten Phase.

Deskriptive Statistik Studienphase 2: TN025							
		Erreichte Profilstufe	**Anzahl MSE**	**Anzahl Wörter**	**Durchschnittliche Länge MSE (MLC)**	**Subordinationsrate pro MSE**	**Rate komplexer NP pro MSE**
N	Gültig	11	11	11	11	11	11
	Fehlend	0	0	0	0	0	0
Summe			254	1563			
Mittelwert		3,727	23,091	142,091	6,243	0,255	0,305
Standardfehler des Mittelwerts		0,195	2,034	11,020	0,249	0,039	0,041

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Deskriptive Statistik Studienphase 2: TN025						
	Erreichte Profilstufe	**Anzahl MSE**	**Anzahl Wörter**	**Durchschnittliche Länge MSE (MLC)**	**Subordinationsrate pro MSE**	**Rate komplexer NP pro MSE**
Median	4,000	21,000	130,000	6,438	0,238	0,316
Modus	4,00	16,00a	103,00	6,44	0,38	0,06a
Std.-Abweichung	0,647	6,745	36,550	0,825	0,128	0,137
Varianz	0,418	45,491	1335,891	0,681	0,016	0,019
Minimum	2	16	103	4,76	0,05	0,06
Maximum	4	38	215	7,50	0,41	0,50
a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.						

Tabelle 57: Deskriptive Daten zu TN025 – zweite Studienphase

Werden die Daten der beiden Teilnehmenden vergleichend gegenübergestellt, ergeben sich zahlreiche Übereinstimmungen zwischen TN024 und TN025. Für beide liegt der Median der erreichten Profilstufen in der ersten Studienphase bei 3 und in der zweiten Studienphase bei 4. Hier zeigt sich, wie bei den Gruppendaten in Kapitel 4.1.2, ein Deckenniveau für die Entwicklung der untersuchten Teilnehmenden bei satzwertigen Einheiten mit Verbendstellung. Gleichzeitig verdeutlicht dies den Fortschritt, den beide in ihrer zweitsprachlichen Entwicklung durchlaufen. Ihre Texte erreichen nicht nur höhere (syntaktische) Profilstufen, sondern sie werden außerdem länger und beinhalten eine größere Anzahl an Nominalphrasen.

Es zeigen sich gleichzeitig Unterschiede in den Daten von TN024 und TN025. Diese werden bei den Variablen ‚Anzahl MSE' und ‚Anzahl Wörter' deutlich. TN025 realisiert in beiden Phasen jeweils mehr MSE als TN024. Zusätzlich schreibt TN025 in beiden Phasen längere Texte, d. h. Texte mit mehr Wörtern als TN024. Dies gilt ebenso für die absoluten Werte der Variable ‚Anzahl Wörter'. Weiterhin fällt auf, dass sich das Längenmaß ‚MLC' zwischen TN024 und TN025 deutlich unterscheidet. Während TN025, wie beschrieben, längere Texte mit mehr MSE schreibt, realisiert TN024 deutlich längere minimale satzwertige Einheiten als TN025. Dass sich aus diesem Komplexitätsmaß nur in begrenztem Maß eine höhere Komplexität der Texte von TN024 ableiten lässt, verdeutlichen die Werte der ‚Subordinationsrate', die für beide Studienphasen bei TN025 höher liegen als bei TN024. Während TN024 längere satzwertige Einheiten produziert, verbindet TN025 diese öfter mit satzwertigen Einheiten, die ein Verb in Endstellung beinhalten. Ein weiterer Unterschied findet sich bei der Integration komplexer Nominalphrasen in die eigenen Texte. TN024 produziert in beiden Studienphasen deutlich mehr komplexe NPs pro MSE als TN025. Der Wert liegt in der ersten Studienphase für TN024 so hoch wie für TN025 in der zweiten Studienphase. Die an dieser Stelle herausgearbeiteten Unterschiede lassen, trotz der sehr großen

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Ähnlichkeiten und Gemeinsamkeiten zwischen den beiden Teilnehmenden, darauf schließen, dass Texte, die sich auf dieselbe Aufgabenstellung beziehen, von TN024 und TN025 in vielen Fällen sehr unterschiedlich ausgearbeitet werden. Dies wird im folgenden Kapitel, vor allem aber in den Kapiteln 4.2.3 und 4.2.4, genauer in den Blick genommen.

Die Betrachtung der deskriptiven Statistiken zeigt einige Tendenzen in den Daten auf. Die Werte aller Variablen nehmen im Studienverlauf zu. Die beiden Teilnehmenden produzieren im Zeitverlauf längere Texte, in denen mehr und längere minimale satzwertige Einheiten verwendet werden. Diese Texte scheinen darüber hinaus in der zweiten Studienphase komplexer als in der ersten zu sein. Dafür sprechen die zunehmenden Werte der Variablen ‚Subordinationsrate pro MSE' und ‚Rate komplexer Nominalphrasen pro MSE'. Es lassen sich Unterschiede zwischen den Teilnehmenden ausmachen, die sich vor allem auf die Länge der minimalen satzwertigen Einheiten und die Rate komplexer Nominalphrasen pro MSE beziehen und die darauf schließen lassen, dass die beiden Teilnehmenden auf dieselben Aufgabenstellungen unter Umständen sehr unterschiedlich reagieren. Dies wird in den nächsten Analyseschritten an einzelnen Texten überprüft, da bisher nur deskriptive und nach Studienphase aggregierte Daten interpretiert werden. In allen Tabellen wird darüber hinaus anhand der Standardabweichungen und der zum Teil weit auseinanderliegenden Werte für das Minimum und das Maximum in den einzelnen Datensets deutlich, dass in den Daten relativ viel Variabilität für jede/-n der beiden Teilnehmenden vorhanden ist. Diese intraindividuelle Variabilität wird im nächsten Kapitel in den Fokus gerückt und mithilfe von min-max-Grafiken sichtbar gemacht.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

TN	Text	Bezeichnung	Datum	Wochen nach Beginn	Besuchte UE	Ermittelte Profil-stufe	Anzahl MSE	Anzahl Wörter	Durch-schnittli-che Länge MSE (MLC)	Anzahl NS	Sub-ordina-tionsrate (DC/C)	Komplexe Nominal-phrasen (CompNP)	Komplexe Nominalphra-sen pro MSE (ComNP/C)
TN024	1	Sich vorstellen	26.05.2016	0	15	1	9	44	4,889	0	0	3	0,333
TN024	2	Was machen Sie nach dem Deutsch-kurs?	28.06.2016	5	98	3	11	68	6,182	0	0	1	0,091
TN024	3	Was hast du ge-macht?	11.07.2016	7	110	3	7	47	6,714	0	0	1	0,143
TN024	4	Weg zum Institut	15.07.2016	7	130	3	10	87	8,700	0	0	1	0,1
TN024	5	Regeln in Deutsch-land	05.08.2016	10	201	3	12	62	5,167	0	0	3	0,25
TN024	6	Mein Tag im Zoo	16.08.2016	12	231	3	13	76	5,846	0	0	6	0,462
TN024	7	A1/A2-Test (Leipzig)	07.09.2016	15	292	3	12	57	4,750	0	0	1	0,083
TN024	8	Ferien	19.09.2016	17	307	3	13	72	5,538	0	0	6	0,462
TN024	9	Das perfekte Abend-essen	30.09.2016	18	347	4	10	59	5,900	3	0,3	4	0,4
TN024	10	Bildergeschichte	31.10.2016	23	440	3	16	87	5,438	2	0,125	11	0,688
TN024	11	A2-Test (Berlin)	14.11.2016	25	481	2	9	60	6,667	0	0	5	0,556
TN024	13	B1-Prüfung	12.12.2016	29	550	1	19	92	4,842	2	0,125	5	0,263
TN024	15	Lieblingskünstler	24.01.2017	35	573	4	18	122	6,778	3	0,167	10	0,556
TN024	16	Karte	08.03.2017	41	668	1	12	112	9,333	1	0,083	11	0,917
TN024	17	Fluchtgeschichte	16.03.2017	42	697	4	28	185	6,607	3	0,107	15	0,536
TN024	18	Lieblingsfilm	03.04.2017	45	756	3	9	62	6,889	1	0,111	4	0,444
TN024	19	Journal	05.04.2017	45	766	4	17	111	6,529	3	0,176	10	0,588
TN024	20	Als ich 18 war	07.04.2017	45	771	4	19	116	6,105	6	0,316	9	0,474
TN024	21	Mein Stadtteil	26.04.2017	48	811	3	17	126	7,412	1	0,059	8	0,471
TN024	22	Ehe auf Zeit	16.05.2017	51	857	3	14	92	6,571	1	0,071	8	0,571
TN024	23	Frühförderung	19.07.2017	60	1010	4	19	134	7,053	6	0,316	7	0,368
TN024	24	Arbeit	28.07.2017	61	1037	4	16	129	8,063	4	0,250	15	0,938

Tabelle 58: Übersicht Daten TN024

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

TN	Text	Bezeichnung	Datum	Wochen nach Beginn	Besuchte UE	Ermittelte Profil-stufe	Anzahl MSE	Anzahl Wörter	Durch-schnittli-che Länge MSE (MLC)	Anzahl NS	Sub-ordina-tionsrate (DC/C)	Komplexe Nominal-phrasen (CompNP)	Komplexe Nominalphra-sen pro MSE (ComNP/C)
TN025	1	Sich vorstellen	26.05.2016	0	15	1	8	37	4,625	0	0	3	0,375
TN025	2	Was machen Sie nach dem Deutsch-kurs?	28.06.2016	5	98	3	10	70	7,000	0	0	3	0,300
TN025	3	Was hast du ge-macht?	11.07.2016	7	110	2	9	49	5,444	0	0	1	0,111
TN025	4	Weg zum Institut	15.07.2016	7	130	1	8	43	5,375	0	0	2	0,250
TN025	5	Regeln in Deutsch-land	05.08.2016	10	205	3	15	107	7,133	0	0	1	0,067
TN025	7	A1/A2-Test (Leipzig)	07.09.2016	15	285	3	8	36	4,500	0	0	1	0,125
TN025	8	Ferien	19.09.2016	17	300	3	15	101	6,733	0	0	6	0,400
TN025	9	Das perfekte Abend-essen	30.09.2016	18	338	4	11	58	5,273	3	0,273	0	0
TN025	10	Bildergeschichte	31.10.2016	23	436	3	34	122	3,588	0	0	4	0,118
TN025	11	A2-Test (Berlin)	14.11.2016	25	486	2	11	63	5,727	2	0,220	6	0,545
TN025	12	Mein schönster Glücksmo-ment	05.12.2016	28	558	3	13	67	5,154	2	0,231	3	0,231
TN025	13	B1-Prüfung	12.12.2016	29	582	4	21	107	5,095	4	0,220	1	0,048
TN025	14	Ferien_2	04.01.2017	32	592	4	16	103	6,438	3	0,188	1	0,063
TN025	15	Lieblingskünstler	24.01.2017	35	632	2	19	111	5,842	1	0,053	6	0,316
TN025	16	Karte	08.03.2017	41	707	3	25	119	4,760	1	0,067	10	0,400
TN025	17	Fluchtgeschichte	16.03.2017	42	734	4	19	130	6,842	4	0,211	9	0,474
TN025	18	Lieblingsfilm	03.04.2017	45	793	4	21	121	5,762	4	0,190	8	0,381
TN025	19	Journal	05.04.2017	45	803	4	20	144	7,200	5	0,238	10	0,500
TN025	20	Als ich 18 war	07.04.2017	45	808	4	16	103	6,438	6	0,375	3	0,188
TN025	21	Mein Stadtteil	26.04.2017	48	855	4	24	163	6,792	6	0,292	8	0,333
TN025	22	Ehe auf Zeit	16.05.2017	51	911	4	32	174	5,438	13	0,406	4	0,125
TN025	23	Frühförderung	19.07.2017	60	1051	4	38	215	5,658	15	0,405	11	0,289
TN025	24	Arbeit	28.07.2017	61	1080	4	24	180	7,500	9	0,375	7	0,292

Tabelle 59: Übersicht Daten TN025

4.2.2 Grafische Analyse: Variabilität

Nach diesem generellen Überblick über die deskriptiven Daten von TN024 und TN025 soll im nächsten Schritt die intraindividuelle Variabilität innerhalb der ausgewählten Variablen sichtbar gemacht und analysiert werden (für eine ausführliche Diskussion des in dieser Arbeit zugrunde gelegten Konzepts vgl. Kapitel 2.5.1 und Kapitel 3.3.3). Variabilität verdeutlicht aus CDST-Sicht die Anpassungsfähigkeit bzw. die Flexibilität eines Systems. Es wird angenommen, dass sie verantwortlich für eine Entwicklung innerhalb des betrachteten Systems sein kann. Gleichzeitig könnte sie einen bestimmten Zeitpunkt innerhalb dieser Entwicklung markieren (vgl. Bot / Lowie / Verspoor 2007). Eine in bestimmten Zeitphasen hohe Variabilität innerhalb eines lernersprachlichen Systems kann zudem als Indikator für die Neuorientierung eines Systems dienen. Geringere Variabilität könnte hingegen als Indikator für relative Stabilität des beobachteten lernersprachlichen Systems gedeutet werden. Übertragen auf zweitsprachliche Lernprozesse könnte Variabilität demnach geringer werden, falls Lernprozesse bestimmte ‚Endpunkte' erreichen (vgl. van Dijk / Verspoor / Lowie 2011: 76).[213]

Die Variabilität wird grafisch mithilfe von min-max-Diagrammen herausgearbeitet (zur Technik dieser Diagramme vgl. Kapitel 3.3.3). In den vorliegenden Abbildungen wird ein bewegliches Fenster (*moving window*) mit dem Intervall ‚5' gewählt. Dieses Fenster wird in jedem Berechnungsschritt um genau einen Zeitpunkt bzw. Text weiterbewegt. Die Graphen für das Minimum und das Maximum sind jeweils ab dem dritten Zeitpunkt angegeben, da dieser den Mittelpunkt des ersten angelegten Intervalls bildet. Im Folgenden werden die min-max-Diagramme für die verschiedenen Wortstellungsmuster und für die übrigen syntaktischen Variablen (‚Erreichte Profilstufe', ‚Durchschnittliche Länge MSE', ‚Subordinationsrate pro MSE' und ‚Rate komplexer Nominalphrasen pro MSE') gezeigt und nacheinander besprochen. Je weiter die minimalen und die maximalen Graphen auseinanderliegen, desto größer ist die angenommene Variabilität in den Daten. Im Gegensatz dazu kann von einer größeren Stabilität bzw. geringeren Variabilität ausgegangen werden, wenn die beiden Graphen näher beieinanderliegen oder in einem gleichbleibenden Abstand zueinanderstehen. Die Ergebnisse dieser grafischen Annäherungen an die Daten werden dann sehr eng auf die Ergebnisse der CPAs bezogen,

213 Anhand der hier besprochenen Variablen sind konkrete Lernprozesse oder ihre angenommenen Endpunkte (wie z. B. der Erwerb der Verbalklammer) sicherlich nur über Umwege abzulesen. Hierfür wären unter Umständen obligatorische Kontexte oder Korrektheitsraten oder andere Variablen, die (implizit oder explizit) eine L1-Norm ansetzen, zielführender, um einen teleologischen Entwicklungsprozess zu verdeutlichen. Eine Ausnahme bildet die Variable ‚Erreichte Profilstufe', die obligatorischen Kontexte einbezieht und daher eine L1-Norm anlegt und zusätzlich eine Verwendungshäufigkeit spezieller Strukturen zur Grundlage ihrer Berechnung hat, aber eine globale (und relativ grobe) Einordnung der Texte markiert. Die identifizierten obligatorischen Kontexte wurden ebenfalls quantifiziert, aber nicht in die vorliegenden Analysen miteinbezogen.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

um eventuelle Entwicklungsschritte zu identifizieren und diese auf mögliche Verbindungen zu in den Daten identifizierter Variabilität zu überprüfen.

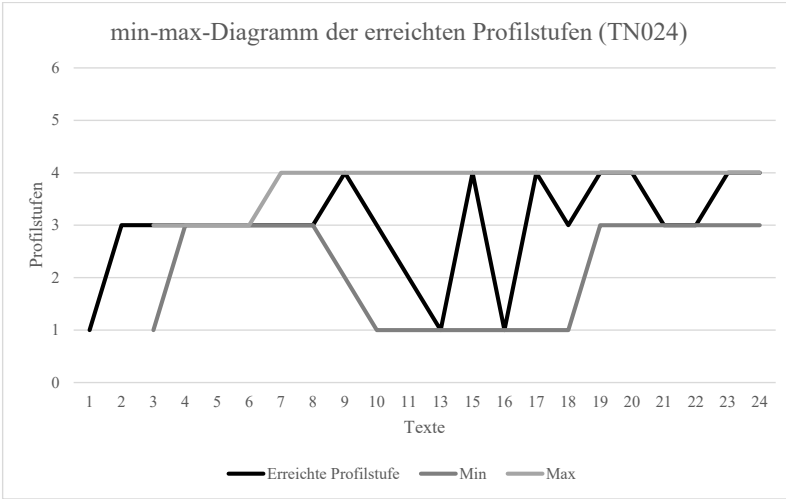


Abbildung 28: min-max Diagramm der erreichten Profilstufen (TN024)

Im min-max-Diagramm für TN024 (vgl. Abbildung 28) lassen sich in Bezug auf die erreichten Profilstufen drei Phasen unterschiedlich großer Variabilität feststellen. Eine erste Phase kann zwischen Zeitpunkt 1 (nach 3 Tagen bzw. in der ersten Unterrichtswoche / nach 15 Unterrichtseinheiten) und Zeitpunkt 6 (nach 12 Wochen / 231 UE) identifiziert werden, in der im Grunde keine Variabilität zu beobachten ist.[214] Die minimale und die maximale Linie liegen aufeinander. Daran schließt sich eine zweite Phase an, die durch eine größer werdende und schließlich konstant groß bleibende Variabilität gekennzeichnet ist und sich von Zeitpunkt 7 (nach 15 Wochen / 292 UE) bis zu Zeitpunkt 18 (nach 45 Wochen / 756) erstreckt. Eine letzte Phase zwischen Zeitpunkt 19 (nach 45 Wochen / 766) und Zeitpunkt 24 (nach 61 Wochen / 1037 UE) ist durch eine geringer werdende Variabilität und damit größere Stabilität gekennzeichnet. Auffallend ist, dass die B1-Prüfung und die Vorbereitung auf diese in die Phase der größten Variabilität fällt.

214 Hier und im Folgenden werden die Begriffe ‚Zeitpunkt' und ‚Text' synonym verwendet. Für eine Übersicht über die einzelnen Texte bzw. Zeitpunkte für TN024 und TN025 vgl. Tabelle 58 und Tabelle 59. In den Klammern werden die Wochen seit Beginn des Sprachkurses und die besuchten Unterrichtsstunden bis zu diesem Zeitpunkt nach dem Muster (nach X Wochen / X UE) angegeben. Hier werden jeweils die von den beiden Teilnehmenden tatsächlich besuchten Unterrichtseinheiten angegeben (vgl. dazu Tabelle 58 und Tabelle 59).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

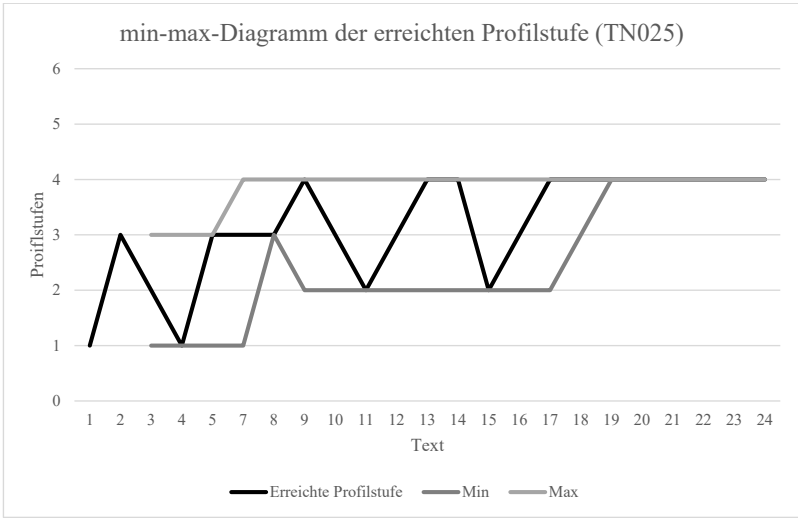


Abbildung 29: min-max Diagramm der erreichten Profilstufen (TN025)

Für TN025 (vgl. Abbildung 29) ergibt sich ein etwas anderes Bild. Es sind ebenfalls mehrere Phasen sich verändernder Variabilität sichtbar. Im Gegensatz zu TN024 lässt sich für TN025 zu Beginn der Studie zwischen Zeitpunkt 1 (nach drei Tagen bzw. in der ersten Woche des Unterrichts / 15) und Zeitpunkt 7 (nach 15 Wochen Unterricht / 285 UE) eine relativ große Variabilität feststellen. Die Ergebnisse schwanken zwischen Profilstufe 1 und Profilstufe 3. Die Variabilität wird am größten zwischen Zeitpunkt 5 (nach zehn Wochen / 205 UE) und Zeitpunkt 7 (der Text zu Zeitpunkt 6 liegt für TN025 nicht vor). Dann scheint es zu einem Sprung in die zweite Phase zu kommen. Die Bandbreite der Variabilität bleibt in dieser Phase konstant. Die Ergebnisse schwanken jetzt zwischen Profilstufe 2 und Profilstufe 4. Diese Phase beginnt zu Zeitpunkt 8 (nach 17 Wochen / 300 UE) und hält bis zu Zeitpunkt 18 (nach 45 Wochen / 793 UE). Die letzte Phase ab Zeitpunkt 19 (nach 45 Wochen / 803 UE) ist gekennzeichnet durch eine größere Stabilität in den beobachteten Daten. Es wird ab Zeitpunkt 17 nur noch Profilstufe 4 erreicht, die damit in dieser Phase der zweitsprachlichen Entwicklung ein Deckenniveau für TN025 darzustellen scheint.

Die zweite Variable, die auf Variabilität geprüft wird, ist die durchschnittliche Länge minimaler satzwertiger Einheiten (MLC), die von TN024 und TN025 in den verschiedenen Texten produziert werden. Für TN024 (vgl. Abbildung 30) lassen sich hier vier Phasen unterscheiden. Zwei Phasen relativ großer Variabilität werden von zwei Phasen größerer Stabilität abgelöst. Die erste Phase größerer Variabilität beginnt zu Zeitpunkt 1 (nach 3 Tagen bzw. in der ersten Unterrichtswoche / 15 UE) und endet mit der stark abnehmenden Variabilität

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

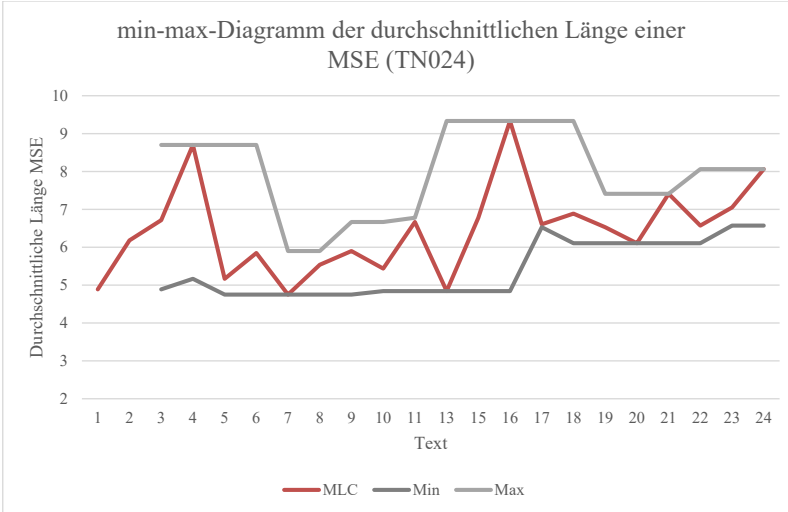


Abbildung 30: min-max Diagramm der durchschnittlichen Länge einer MSE (TN024)

zu Zeitpunkt 6 (nach zwölf Wochen / 231 UE). Die nächste Phase, die als eher variabel beschrieben werden kann, beginnt zu Zeitpunkt 11 (nach 25 Wochen / 481 UE) und endet zu Zeitpunkt 18 (nach 45 Wochen / 756 UE). Zwischen Zeitpunkt 7 (nach 15 Wochen / 292 UE) und 11 (nach 25 Wochen / 481 UE) und zwischen Zeitpunkt 19 (nach 45 Wochen / 766 UE) und 24 (nach 61 Wochen / UE 1037 UE) bleiben die Bandbreiten der beobachteten Variabilität sowohl relativ gering als auch sehr stabil. Sehr deutlich treten zwei Zeitpunkte hervor, zu denen im Vergleich zu den restlichen Texten deutlich längere satzwertige Einheiten produziert werden. Dies ist in Text 4 und in Text 16 der Fall. Diese beiden Zeitpunkte beeinflussen damit in diesem methodischen Ansatz stark die sichtbare Variabilität.

Im Gegensatz zu TN024 wirkt das min-max-Diagramm der durchschnittlichen Länge einer MSE für TN025 (vgl. Abbildung 31) deutlich stabiler und in den Daten finden sich keine deutlich sichtbaren Ausreißer nach oben oder nach unten. Es lassen sich kaum unterscheidbare Phasen in der Variabilität ausmachen. Fast durchgehend liegen die Werte zwischen 4 bzw. 5 und 7. Nur zwischen den Zeitpunkten 7 (nach 15 Wochen / 285 UE) und 12 (nach 28 Wochen / UE 558 UE) sinken sowohl die minimale als auch die maximale Grenze ein wenig nach unten ab und die Variabilität wird etwas größer. Deutlich wird ebenfalls, dass die durchschnittliche Länge einer MSE im Zeitverlauf der Datenerhebung nicht wesentlich zunimmt. Zum Teil werden in den Texten zu Beginn der Studie längere MSE als am Ende der Studie produziert.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

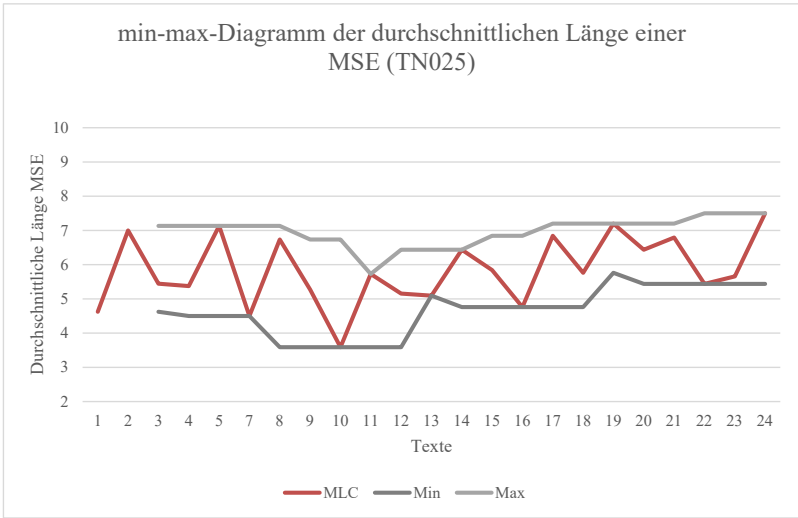


Abbildung 31: min-max Diagramm der durchschnittlichen Länge einer MSE (TN025)

Wie im Kapitel zu den deskriptiven Statistiken der Daten von TN024 und TN025 angedeutet, werden die ersten Nebensatzstrukturen erst ab Text Nummer 8 (nach 17 Wochen / 307 UE) produziert. Dies wirkt sich ebenfalls recht deutlich auf die beiden min-max-Diagramme zur Subordinationsrate bzw. auf den Anteil der unterordnenden Nebensätze an den insgesamt produzierten MSE aus. Wird diese erste Phase aus der Analyse ausgeklammert, ergeben sich für TN024 (vgl. Abbildung 32) drei unterschiedliche Phasen in Bezug auf die Variabilität der Variable ‚Subordinationsrate'.[215] Eine erste Phase, die durch höhere Variabilität gekennzeichnet ist, beginnt ab dem Zeitpunkt der ersten Produktion von Nebensätzen und dauert an bis zu Zeitpunkt bzw. Text 13 (nach 29 Wochen / 550 UE). Daraufhin verringert sich die Variabilität und stabilisiert sich ein wenig zwischen Zeitpunkt 15 (nach 35 Wochen / 573 UE) und Zeitpunkt 17 (nach 42 Wochen / 697 UE), um dann ab Zeitpunkt 18 (nach 45 Wochen / 756 UE) bis zum Ende der Studie wieder deutlich zuzunehmen. Dabei überschreiten die Werte der Subordinationsrate nie das Niveau von 0,3. Maximal ist jede dritte satzwertige Einheit ein Nebensatz mit Verbendstellung. Dieser Wert wird zu Zeitpunkt 9, dem ersten Zeitpunkt, an dem Nebensatzstrukturen produziert werden, erreicht. Dies könnte ein Indikator für einen gewissen Over-Use sein, der mit der Kenntnis einer neu im Unterricht eingeführten Struktur einhergeht. In Text 9

215 Zu diskutieren wäre an dieser Stelle, ob es sich im Falle einer noch nicht emergenten Struktur oder bei Komplexitätsmaßen, die, wie in diesem Fall, noch bei null liegen, um eine Phase handelt, die durch große Stabilität gekennzeichnet ist.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

(‚Das perfekte Abendessen') müssen die Lernenden beschreiben, was für sie ein ‚perfektes' Abendessen ausmachen würde und warum.

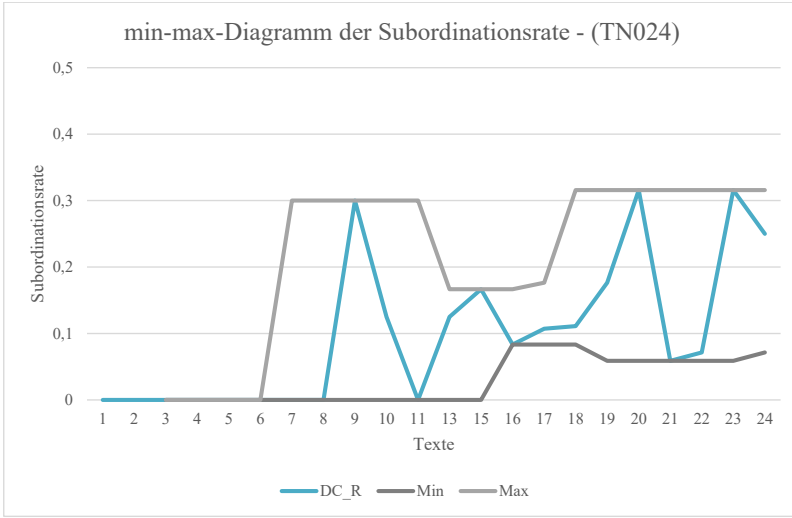


Abbildung 32: min-max Diagramm der Subordinationsrate (TN024)

Wie bei TN024 lässt sich für TN025 (vgl. Abbildung 33) hinsichtlich der Subordinationsrate zunächst eine Phase erkennen, die durch hohe Variabilität gekennzeichnet ist. Diese dauert deutlich länger als bei TN024, nämlich von Zeitpunkt 8 (nach 17 Wochen / 300 UE) bis zu Zeitpunkt 17 (nach 42 Wochen / 734 UE), wobei sich die Variabilität ab Zeitpunkt 12 (nach 28 Wochen / 558 UE) etwas verringert. Ab Zeitpunkt 17 bleibt die Variabilität insgesamt bis zum Ende der Studie geringer als bis zu diesem Zeitpunkt. Gleichzeitig bilden die minimale und die maximale Grenze eine Art steigenden Kanal.[216] In diesem steigen die Werte der Subordinationsrate kontinuierlich an und erreichen die höchsten Werte der gesamten Studiendauer. Gerade zum Ende der Studie liegen die Werte von TN025 deutlich über den Werten von TN024, wobei sie den Wert von 0,4 nicht übersteigen. Es sind folglich maximal 40 % der produzierten MSE Nebensätze mit Verbendstellung. Analog zu TN024 erreicht TN025 hier zu Zeitpunkt 9 innerhalb der Zeitreihe einen sehr hohen Wert, obwohl dies der erste Zeitpunkt ist, an dem Nebensatzstrukturen produziert werden.

216 Damit ist an dieser Stelle der Umstand gemeint, dass der Abstand zwischen dem minimalen und dem maximalen Graphen konstant bleibt, während beide kontinuierlich ansteigen,

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

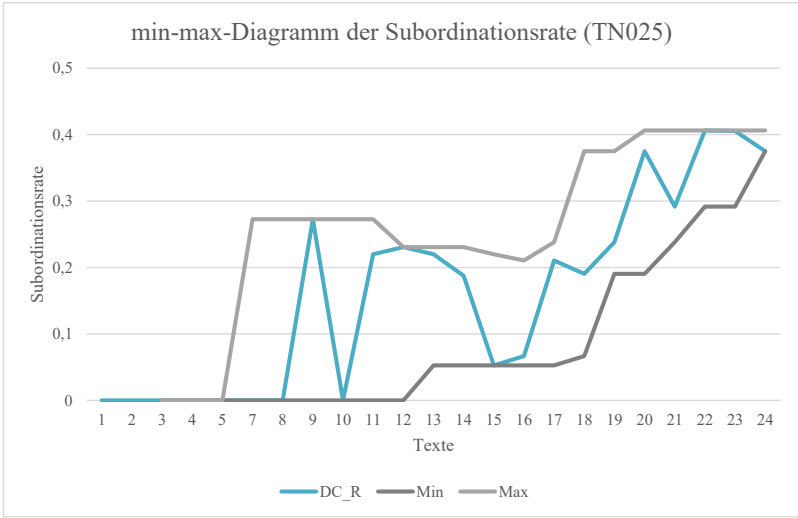


Abbildung 33: min-max Diagramm der Subordinationsrate (TN025)

Die letzte der in Bezug auf die intraindividuelle Variabilität besprochenen Variablen ist die Rate komplexer Nominalphrasen pro MSE. Es lassen sich mehrere Phasen in den Daten von TN024 (vgl. Abbildung 34) identifizieren. Dabei ergibt sich eine Art Stufenform, wobei sich die erste Stufe von Zeitpunkt 3 (nach 7 Wochen / 110 UE) bis Zeitpunkt 7 (nach 15 Wochen / 292 UE) zu erstrecken scheint. Die zweite Stufe reicht von Zeitpunkt 8 (17 Wochen / 307 UE) bis zu Zeitpunkt 11 (nach 25 Wochen / 481 UE) und die dritte Stufe beginnt etwa zu Zeitpunkt 13 (nach 29 Wochen / 550 UE) und endet bei Zeitpunkt 18 (nach 45 Wochen / 756 UE). Die Stufen zeichnen sich jeweils durch ein Plateau aus, in dem sich der Wert für den maximalen Graphen über mehrere Zeitpunkte nicht ändert. Zusätzlich relevant ist, dass sich der minimale und der maximale Graph weitgehend parallel entwickeln. Kommt es etwa zu einem ‚Stufenschritt' im maximalen Graphen, so folgt der minimale Graph in der Regel zwei Zeitpunkte später diesem Muster. Ab Zeitpunkt 17 (nach 42 Wochen / 697 UE) bleibt der min-Graph relativ stabil, während der max-Graph ab Zeitpunkt 18 zunächst stark abnimmt, was auf eine geringere Variabilität in den Daten zwischen Zeitpunkt 19 (nach 45 Wochen / 766 UE) und Zeitpunkt 21 (nach 48 Wochen / 811 UE) schließen lässt. Ab Zeitpunkt 21 steigt der max-Graph wieder an und verdeutlicht eine zum Ende der Studie zunehmende Variabilität in der Rate komplexer Nominalphrasen pro MSE für TN024.

Beim Vergleich der beiden Teilnehmenden fällt für die Rate komplexer Nominalphrasen auf, dass sich die Datenreihe für TN025 (vgl. Abbildung 35) durch größere Stabilität auszeichnet. Es lassen sich schwieriger als bei anderen Daten-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

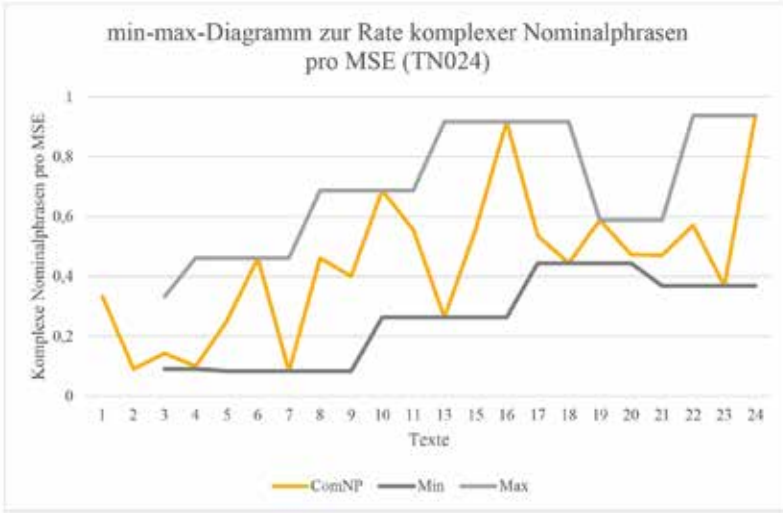


Abbildung 34: min-max Diagramm zur Rate komplexer Nominalphrasen pro MSE (TN024)

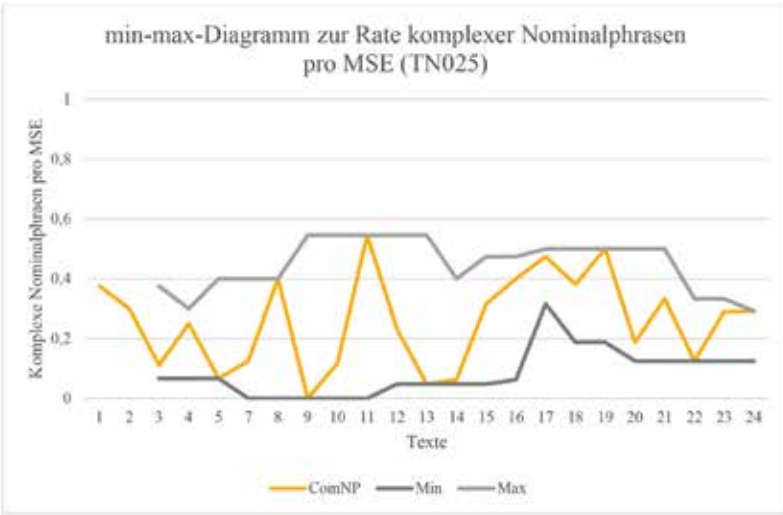


Abbildung 35: min-max Diagramm zur Rate komplexer Nominalphrasen pro MSE (TN025)

reihen unterschiedliche Phasen erkennen. Deutlich wird, dass ab Zeitpunkt 8 (nach 17 Wochen / 300 UE) die Datenreihe durch eine sehr große Variabilität gekennzeichnet ist. Erst zum Ende der Studie, ab Zeitpunkt 17 (nach 42 Wo-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

chen / 734 UE), verringert sich die Bandbreite zwischen den beiden Graphen für TN025, während sie für TN024 zum Ende der Studie stark ansteigt. Die größte Variabilität innerhalb der Daten lässt sich zwischen Zeitpunkt 9 (nach 18 Wochen / 338 UE) und Zeitpunkt 13 (nach 29 Wochen / 582 UE) beobachten.

Die min-max-Grafiken machen für die hier betrachteten Zeitreihen beider Teilnehmender zum Teil sehr unterschiedliche Phasen sich verändernder Variabilität sichtbar. Die Daten von TN024 scheinen insgesamt, mit Ausnahme der Daten zur Subordinationsrate, etwas variablere Phasen als die Daten von TN025 aufzuweisen. Die Daten von TN025 zeichnen sich im Studienverlauf für einige der untersuchten Variablen, etwa für die durchschnittliche Länge minimaler satzwertiger Einheiten, durch größere Stabilität aus. Dies ergibt zusätzlich im Kontext der zuvor auf Gruppenebene durchgeführten Clusteranalysen Sinn. Beide Teilnehmende werden in eher variable Cluster eingeordnet: TN024 in das ‚variabelste ' Cluster 7 und TN025 in das ebenfalls sehr variable Cluster 6 (vgl. für die Einteilung in verschiedene Cluster nach der in den Daten gefundenen Variabilität Kapitel 4.1.3). Gleichzeitig wird sehr deutlich, wie stark sich selbst Teilnehmende unterscheiden, die in eher ähnlichen und in diesem Fall durch größere Variabilität geprägten Clustern eingeordnet wurden.

4.2.3 Grafische Trendanalyse der Interaktionen: Wortstellungsmuster

In diesem Unterkapitel werden die Interaktionen zwischen den verschiedenen Variablen im Kontext eines komplex-dynamischen Ansatzes analysiert. Als Erstes wird die Interaktion der verschiedenen Wortstellungsmuster, die im Rahmen der Profilanalyse ermittelt werden, in den Blick genommen. In jedem Schritt werden zunächst Rohwerte analysiert und dann, um generelle Trends in den Daten sichtbar zu machen, durch einen gleitenden Durchschnitt geglättete Daten analysiert (vgl. für diese Art der Datenvisualisierung Verspoor / van Dijk 2011). Dabei werden für TN024 und TN025 jeweils relative und absolute Werte verglichen, um ein möglichst genaues Bild der Interaktion der verschiedenen syntaktischen Variablen zu erhalten. Dieses Vorgehen hängt damit zusammen, dass die einzelnen Wortstellungsmuster sich teilweise additiv zueinander verhalten. Eine minimale satzwertige Einheit kann immer nur einem Wortstellungsmuster zugeordnet werden und schließt daher die anderen Wortstellungsmuster aus (zur Darstellung komplementärer Variablen mithilfe von Prozentwerten vgl. Verspoor / Schmid / Xu 2012: 245). Ein Satz, der auf Profilstufe 3 eingeordnet wird, könnte durchaus eine korrekt realisierte Separation bzw. eine Struktur der darunterliegenden Profilstufe 2 beinhalten (vgl. dazu Belege [32], [33], und [34] in Kapitel 4.1.2). Die Profilstufe 0 ‚Bruchstücke' wird an dieser Stelle weder in die Analyse miteinbezogen noch in den Schaubildern dargestellt, da es sich bei den Strukturen auf dieser Stufe in den meisten Fällen um Gruß- oder Abschiedsformeln handelt, die daher nur sehr eingeschränkt

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Informationen über die Entwicklung von in der vorliegenden Studie relevanten Wortstellungsmustern in den Texten der beiden Lernenden enthalten. Da TN024 in den untersuchten Texten keine satzwertigen Einheiten der Stufen 5 ‚Integrierter Nebensatz' und 6 ‚Erweitertes Partizipialattribut' produziert, finden sich hierzu in den Analysen ebenfalls keine Daten. TN025 produziert in wenigen Texten eingeschobene Nebensätze (Profilstufe 5), weshalb diese in die Übersicht mitaufgenommen werden. Dies sind zu keinem Zeitpunkt mindestens drei Vorkommen, so dass kein Gesamttext auf dieser Profilstufe eingeordnet wird. Für Profilstufe 6 finden sich in den Texten von TN025 keine Belege. Zuletzt werden jeweils für beide Lernenden die drei syntaktischen Komplexitätsmaße ‚Durchschnittliche Länge einer MSE', ‚Subordination' und ‚Komplexe Nominalphrasen' analysiert.

In den folgenden Schaubildern werden auf der x-Achse die von den beiden Teilnehmenden jeweils produzierten Texte zu den einzelnen Zeitpunkten abgebildet (zu den fehlenden Datenpunkten vgl. Kapitel 3.3.2). Wie im vorangegangenen Kapitel werden die Begriffe ‚Text' und ‚Zeitpunkt' synonym verwendet, da zu jedem Zeitpunkt jeweils maximal nur ein Text für jeden der beiden Teilnehmenden vorliegt. Auf der y-Achse wird die Anzahl der produzierten Wortstellungsmuster abgetragen. Dabei ist die Skalierung der y-Achse an der Maximalzahl der für eines der untersuchten Wortstellungsmuster innerhalb der gesamten Studiendauer produzierten Einheiten orientiert. Dieser Wert liegt bei 18 (TN025, Text 10, 18 satzwertige Einheiten der Profilstufe 1 ‚Finitum'). Für die Schaubilder mit gleitenden Durchschnitten, die sich auf die absoluten Werte beziehen, verschiebt sich der maximale Wert der y-Achse leicht nach unten auf 14. Die y-Achsen zu den relativen Werten sind zwischen den Werten 0 und 1 skaliert, da es sich bei den relativen Werten um prozentuale Anteile handelt. Die folgenden Beschreibungen bzw. Analysen der gezeigten Diagramme können keinen Anspruch auf eine vollständige Beschreibung aller unter Umständen vorhandenen Interaktionen in den Daten erheben. Vielmehr wird versucht, eindrückliche und relevante Interaktionen erklärend hervorzuheben.

Abbildung 36 verdeutlicht, dass TN024 in einem sehr frühen zweitsprachlichen Entwicklungsstadium Sätze mit Inversionen (blaue Linie) realisiert, und zwar zu den meisten Zeitpunkten häufiger als Sätze, in denen ausschließlich eine Separation der Verbteile (rote Linie) beobachtet werden kann. Dies ist unter anderem im zweiten Text (nach 5 Wochen / 98 UE) der Fall. Sätze mit Verbendstellung (grüne Linie) produziert TN024 ab dem neunten Text ‚Das perfekte Abendessen' (nach 18 Wochen / 347 UE). In den Textbeispielen von TN024 finden sich über die gesamte Studiendauer keine Beispiele für Wortstellungsmuster auf den Profilstufen 5 (Integrierte Nebensätze) und 6 (Erweiterte Partizipialattribute). Wird Grießhabers Kriterium von drei Vorkommen pro Text als Erwerbskriterium miteinbezogen, wird deutlich, dass für diesen TN Inversionen (und Nebensätze) zu einem früheren Zeitpunkt als erworben nachgewiesen werden könnten als Satzteile mit Separationen. Vor allem die beiden sehr hohen Werte zu den

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

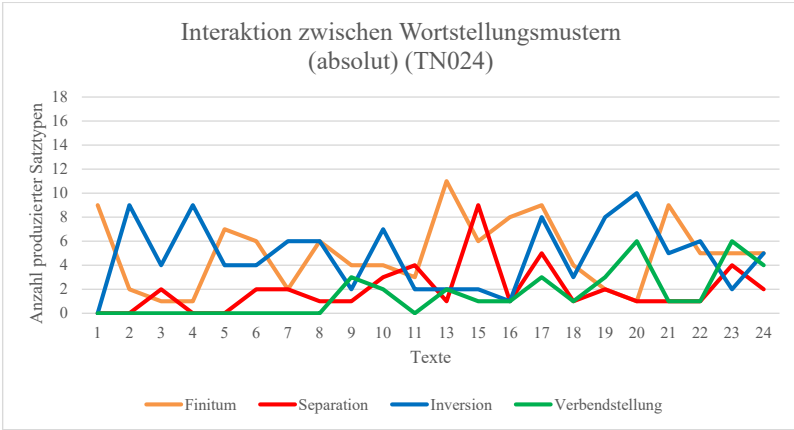


Abbildung 36: Interaktion zwischen den produzierten Wortstellungsmustern für TN024. Gezeigt wird jeweils die absolute Anzahl der produzierten Wortstellungsmuster pro Text (N=22).

Zeitpunkten 2 und 4 (7 Wochen / 130 UE) müssen in diesem Kontext genauer in den Blick genommen werden. Die Abbildung zeigt, dass nur zu einem Zeitpunkt während der Studie höhere Werte als zu diesen beiden initialen Zeitpunkten für die Variable ‚Inversion' vorliegen. Dies ist in Text 20 ‚Als ich 18 war' der Fall, den TN024 nach 45 Wochen bzw. 771 Unterrichtseinheiten produziert. Zu Zeitpunkt 2 ‚Was machen Sie nach dem Deutschkurs?' werden fast ausschließlich untereinander sehr ähnliche Inversionsstrukturen mit adverbialen Ergänzungen im Vorfeld produziert (vgl. Beleg [35]; Inversionsstrukturen mit adverbialen Ergänzungen im Vorfeld sind fettgedruckt):

> [35] ***Heute*** *gehe ich nach dem Deutschkurs zur nach Hause.* ***Dann*** *gehe ich ins Badezimmer. Ich dusche.* ***Dann*** *gehe ich zur Sosialamt. Ich habe einen Termin.* ***Danach*** *kaufe ich im Supermarkt ein.* ***Zu Hause*** *koche ich das Iftaressen.* ***Danach*** *räume ich das Wohnzimmer auf.* ***Am Wochenende*** *gehe ich am Nachmittag zum See.* ***Letzte Woche*** *ging ich ins Konzert Klasikair,* ***aber in dieser Woche*** *werde ich zu Bibliothek gehen.*
>
> (TN024, ‚Was machen Sie nach dem Deutschkurs?')

TN024 verwendet zwar dieselbe Struktur, variiert diese aber inhaltlich deutlich und demonstriert eine zu diesem Zeitpunkt durchaus beeindruckende Bandbreite an ähnlichen adverbialen Ergänzungen im Vorfeld. Es ist zumindest fraglich, ob die hier produzierten Strukturen als Chunks eingestuft werden könnten. In Text 4 ‚Mein Weg zum Institut' soll der eigene Weg zum Max-Planck-Institut, in dem die Sprachkurse stattfanden, beschrieben werden. Diese Aufgabenstellung begünstigt, wie die Aufgabenstellung zu Text 2, die Verwendung von Inversionsstrukturen. Hier produziert TN024 tatsächlich fast ausschließlich satzwertige Einheiten nach dem Muster *Dann gehe ich* […]. *Dann gehe ich* […]. *Dann*

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

gehe ich […] (‚Mein Weg zum Institut, TN024). Selbst wenn diese beiden Texte etwas vorsichtiger als Belege für ein Erreichen der Profilstufe 3 ‚Inversion' gesehen werden sollten, produziert TN024 in den Texten 3+5 und dann in den Texten 6–10 jeweils mehr als drei Inversionsstrukturen und kaum ‚eigenständige' Separationsstrukturen, die die darunterliegende Profilstufe ausmachen. Die teilnehmende Person produziert zum ersten Mal in Text Nummer 10 (nach 23 Wochen) drei Strukturen, die (ausschließlich) als Separation eingestuft werden. An dieser Stelle muss erneut darauf hingewiesen werden, dass Strukturen, die als Inversion gekennzeichnet sind, durchaus klammerbildende Strukturen enthalten (können). Dies ist eine methodische Schwachstelle der Profilanalyse, da dies die Daten unter Umständen verzerren könnte. Die Profilanalyse geht von einer implikationellen Reihenfolge der Stufen aus (vgl. Kapitel 2.4.3). Obwohl Strukturen der Stufe 3 zwar Separationen der Verbteile enthalten können, ist dies keinesfalls zwangsläufig der Fall. Im vorliegenden Fall etwa enthalten vier der neun satzwertigen Einheiten aus dem Text zu Zeitpunkt 2, die in Beispiel a als Inversion gekennzeichnet wurden, Klammerstrukturen. Für die anderen fünf Belege ist dies nicht der Fall. Beide Strukturen werden in den Texten von TN024 folglich zum selben Zeitpunkt emergent.[217] Zu Zeitpunkt 11 beispielsweise wurde die darüberliegende Stufe ‚Inversion' siebenmal und die Stufe ‚Verbendstellung im Nebensatz' einmal erreicht bzw. im Text als erreicht markiert. Dennoch ist bemerkenswert, dass TN024 bereits in einer sehr frühen Phase der zweitsprachlichen Entwicklung frequent Strukturen auf einer höheren Profilstufe produziert. Dies könnte ein Hinweis auf die unter Umständen eingeschränkte Validität eines Verfahrens wie der Profilanalyse für erwachsene Lernende in einem gesteuerten Lernprozess sein, da sie als Analyseinstrument vielleicht nicht genau genug zwischen dem Erwerb verschiedener Stufen differenzieren kann.

Eine andere mögliche Einschränkung der Aussagekraft einer Profilanalyse wird ebenfalls sichtbar. Falls etwa eine einmalige Profilanalyse zu Zeitpunkt 13 (nach 29 Wochen / 550 UE) durchgeführt würde (das ist die B1-Prüfung), wäre das Ergebnis dieser Profilanalyse, dass Stufe 1 und Stufe 2 erreicht wären, aber Stufe 3, die davor kontinuierlich produziert wird, und Stufe 4, die im Text davor erreicht wird, noch nicht als erworben gelten können. Hier würde die Profilanalyse im Gegensatz zu den longitudinalen Daten stark abweichende Ergebnisse liefern. Aus longitudinaler Perspektive zeigt sich, dass TN024 zu keinem Zeitpunkt innerhalb der Studie mehr Strukturen auf Profilstufe 1 produziert als zu Zeitpunkt 13 und es sich hierbei durchaus um eine gezielte Strategie im Prüfungskontext handeln könnte, bei der u. U. komplexere Strukturen (oder längere voneinander abhängige Satzstrukturen) bewusst vermieden werden, um möglichst ‚korrekt' zu schreiben. Deutlich wird diese Spitze in der Produktion

217 Eine simultane Emergenz mehrerer Strukturen wird in der *Processability Theory* auf theoretischer Ebene ausgeschlossen. Ein solcher Fall könnte nur damit erklärt werden, dass die niedrigere Stufe bereits zu einem früheren Zeitpunkt emergiert sein muss (vgl. Kapitel 2.4.1).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

von satzwertigen Einheiten auf Profilstufe 1 im Vergleich zu den anderen Profilstufen, wenn die prozentualen Anteile der einzelnen Wortstellungsmuster verglichen werden (vgl. dazu Abbildung 38). In den direkt auf die B1-Prüfung folgenden Texten werden die übrigen Wortstellungsmuster wieder mit einer deutlich höheren Frequenz produziert und der Anteil an Strukturen auf Profilstufe 1 nimmt ab. Auffällig ist darüber hinaus der Spitzenwert für Profilstufe 2 ‚Separation', der zu Zeitpunkt 15 (nach 35 Wochen / 573 UE) erreicht wird. In diesem Text soll der eigene Lieblingskünstler beschrieben werden. Interessanterweise erreicht TN025 in diesem Text ebenso einen Höchstwert für satzwertige Einheiten auf Profilstufe 2 (vgl. dazu Abbildung 39).

Um die den Daten zugrundeliegenden Trends besser nachvollziehen zu können, werden im Folgenden dieselben Daten für TN024 mithilfe eines gleitenden Durchschnitts (*Moving Average*) geglättet. Dabei sei daran erinnert, dass geglättete Daten jeweils mehrere Datenpunkte gleichzeitig einbeziehen und daher keine Informationen über die exakten Zeitpunkte, an denen Veränderungen stattfinden, liefern können.

In Abbildung 37 sind die durch einen *Moving Average* bzw. einen gleitenden Durchschnitt geglätteten Interaktionsdaten der produzierten Wortstellungsmuster für TN024 zu sehen (vgl. Kapitel 3.3.3; vgl. zu dieser Art von Visualisierung Verspoor / van Dijk (2011)). Ein gleitender Durchschnitt ermöglicht es, Trends direkter nachzuvollziehen, da nicht einzelne Messpunkte isoliert werden, sondern immer ein Messbereich mit mehreren Messpunkten gleichzeitig betrachtet wird. Dadurch werden die ursprünglichen Daten geglättet und erscheinen nicht mehr so ‚spitz'. In den vorliegenden Abbildungen wurde jeweils ein Messbereich von drei Messpunkten bzw. Texten gewählt. Dies führt unter anderem dazu, dass ab dem sechsten Zeitpunkt der Wert für die Sätze mit Verbendstellung leicht ansteigt, obwohl diese Struktur erst ab Zeitpunkt 8 produziert wird. Durch die Anwendung des gleitenden Durchschnitts spielt diese Struktur ab dem Zeitpunkt eine Rolle, ab dem der Wert von Zeitpunkt 8 in das gleitende Fenster miteinbezogen wird. Bei einem Fenster von drei Messpunkten ist dies ab Zeitpunkt 6 der Fall. In der Beschreibung der Abbildungen wird im Folgenden dennoch die Beschreibung ‚Zeitpunkt / Text' verwendet, um Phasen einzugrenzen.

Vor allem die Wortstellungsmuster ‚Finitum' und ‚Inversion' scheinen in einem kompetitiven Verhältnis zueinanderzustehen. Dies zeigt sich zwischen den Datenpunkten zu Text 1 und Text 4 oder zwischen Text 9 und Text 21. Wenn der Wert der einen Variable steigt, fällt der Wert der anderen Variable und *vice versa*. Es zeigen sich ebenso längere Phasen, in denen das Verhältnis der beiden Wortstellungsmuster konstant ist und sie sich gemeinsam entwickeln, etwa zwischen Text 5 und Text 9. In dieser Phase nehmen die Werte beider Variablen gemeinsam ab. Ebenso relevant ist die Beziehung der Wortstellungsmuster ‚Separation' und ‚Inversion', deren Beziehung über weite Strecken der Studiendauer eher kompetitiv erscheint. Dies lässt sich etwa zwischen Text 16 und Text 23

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

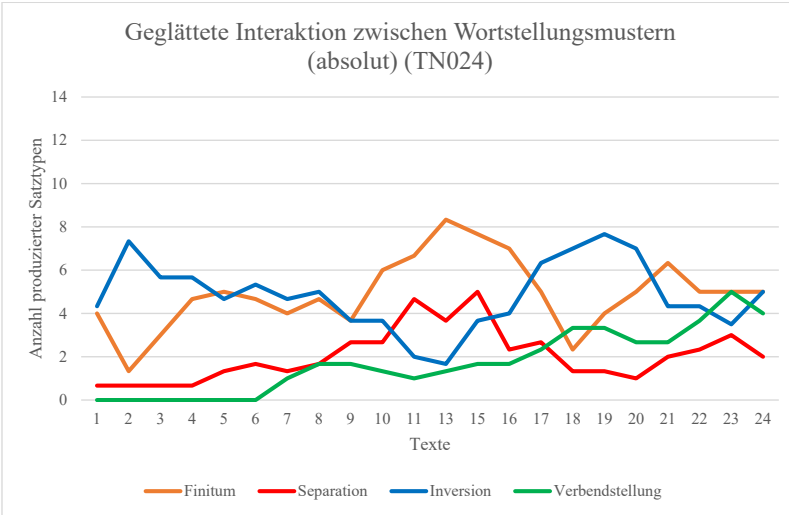


Abbildung 37: Durch Moving Average geglättete Interaktion zwischen den produzierten Wortstellungsmustern für TN024. Der Moving Average hat ein Intervall von 3 und wurde auf der Grundlage der absoluten Anzahl der jeweils produzierten Wortstellungsmuster berechnet.

beobachten. Zusätzlich fällt eine Phase ähnlicher Entwicklung von Zeitpunkt 8 bis Zeitpunkt 16 auf. In dieser Phase nähern sich die beiden Wortstellungsmuster sich in ihrer Frequenz an, um dann ab Zeitpunkt 16 wieder die beschriebene kompetitive Phase einzutreten. Der Beginn dieser gemeinsamen Entwicklungsphase fällt mit dem Anstieg von MSE mit Verbendstellung zusammen. Hier könnte es sich um eine konditionelle Entwicklung im Zusammenspiel dreier Variablen handeln.

Deutlich wird in Abbildung 37 und an der Darstellung der prozentualen Anteile (vgl. Abbildung 38), dass die Variabilität zwischen den Wortstellungsmustern zum Ende der Studie hin abnimmt. Mit geringer werdender Variabilität scheint sich ein größeres Gleichgewicht zwischen den verschiedenen Wortstellungsmustern einzustellen. Dies könnte durchaus ein Indiz dafür sein, dass die teilnehmende Person größere Sicherheit im Umgang mit verschiedenen Satztypen erreicht hat und es nicht mehr zur übermäßigen Nutzung einzelner Muster kommt. Solche sehr stabilen Zustände des lernersprachlichen Systems werden in der Theoriebildung der CDST als Equilibrium bezeichnet (vgl. Kapitel 2.5.1).

Noch eindeutiger wird die Beziehung der beiden Wortstellungsmuster ‚Finitum' und ‚Inversion' für TN024, wenn anstatt der absoluten Zahlen der relative bzw. prozentuale Anteil der produzierten Wortstellungsmuster in jedem Text in den Blick genommen wird. Wie in Abbildung 37 wird in Abbildung 38 ein

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

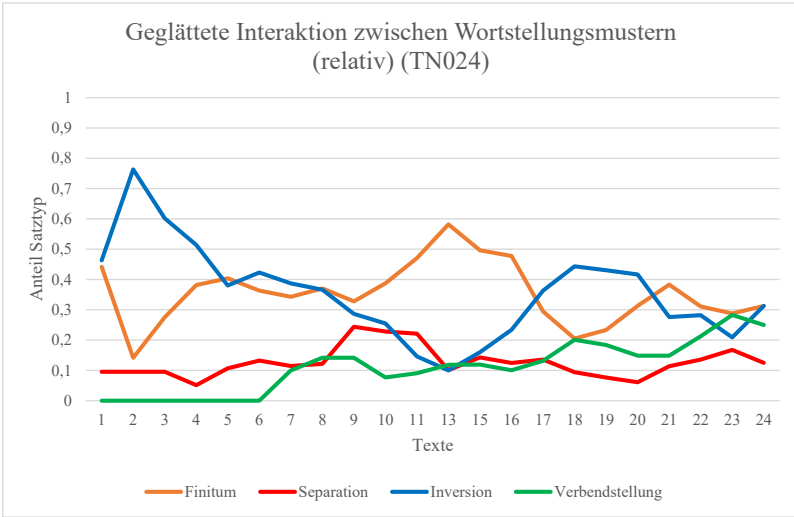


Abbildung 38: Durch Moving Average geglättete Interaktion zwischen den produzierten Wortstellungsmustern für TN024. Der Moving Average hat ein Intervall von 3 und wurde auf der Grundlage der relativen Anzahl der jeweils produzierten Wortstellungsmuster berechnet. Gezeigt werden die prozentuellen Anteile der in den Texten produzierten Wortstellungsmuster.

gleitender Durchschnitt mit dem Intervall 3 zur Glättung der Daten eingesetzt. Immer wenn die Sätze mit kanonischem Wortstellungsmuster (‚Finitum') einen hohen Anteil innerhalb der produzierten Texte aufweisen, haben die Sätze mit ‚Inversion' nur einen geringen Anteil und andersherum. Bei der überwiegenden Mehrzahl der Strukturen auf Profilstufe 3 handelt es sich um satzwertige Einheiten mit adverbialen Ergänzungen im Vorfeld, die durchaus als Hinweise auf den Einsatz von komplexeren textverknüpfenden Strukturen gelesen werden könnten. In diesem Zusammenhang wird erneut deutlich, dass TN024 vor allem im Kontext der B1-Prüfung verstärkt auf weniger komplexe Strukturen zurückgreift.

Noch deutlicher als in Abbildung 37 wird hier, dass der Anteil der Sätze, die ausschließlich als Separation klassifiziert werden, für TN024 über die gesamte Studiendauer relativ stabil bleibt. Ebenfalls deutlich zu sehen ist der relativ stabile Zustand, der sich in den letzten Texten der Studie (etwa ab Text 17) einzustellen scheint. In diesen macht keines der Wortstellungsmuster anteilig mehr als 30 % aus.

Im Folgenden werden die Daten für TN025 mithilfe desselben analytischen Dreischritts in den Blick genommen.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

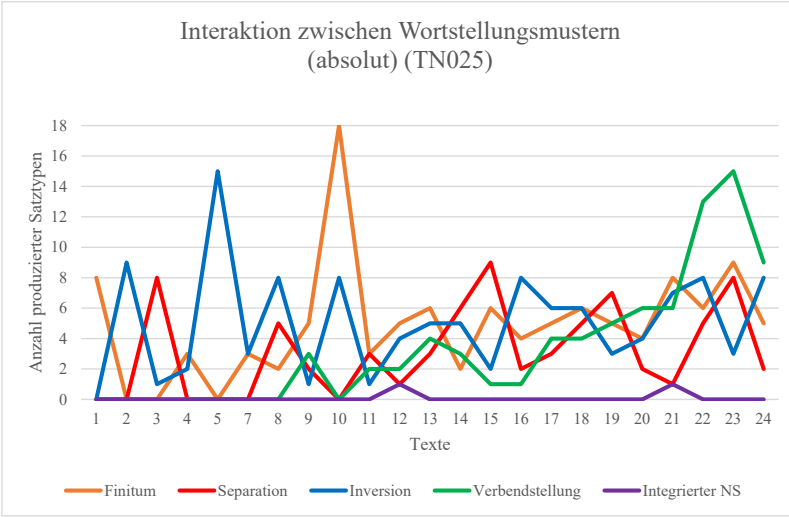


Abbildung 39: Interaktion zwischen den produzierten Wortstellungsmustern für TN025. Gezeigt wird jeweils die absolute Anzahl der produzierten Wortstellungsmuster pro Text (N=23).

Abbildung 39 zeigt die Interaktion der verschiedenen Wortstellungsmuster für TN025 anhand der absoluten Werte. Einige Tendenzen werden deutlich. Wie bei TN024 fällt für TN025 die frühe und frequente Nutzung von satzwertigen Einheiten mit Inversionsstrukturen auf (blaue Linie). Ebenso fällt die größere Variabilität in der Verwendung der verschiedenen Wortstellungsmuster auf. Immer wieder gibt es Texte, in denen ein bestimmtes Wortstellungsmuster sehr dominant produziert wird. In Text 5 ‚Regeln in Deutschland' (nach zehn Wochen / 205 UE) werden ausschließlich Inversionsstrukturen produziert. Diese folgen alle dem Muster *In XX darf / soll / muss man nicht*. Hier kann im Vergleich zu Text 2 von TN024 eher von einer musterartigen Textproduktion ausgegangen werden. In Text 10 (23 Wochen / 436 UE) werden nur Strukturen der beiden Stufen 1 und 3 produziert. Gleichzeitig stellt der Wert von 18 Strukturen der Profilstufe ‚Finitum' den höchsten Einzelwert für ein produziertes Wortstellungsmuster dar, der in den Daten von TN025 zu finden ist (orangene Linie).

Insgesamt scheint es in der ersten Studienphase bis zu Text 13 (B1-Prüfung, 29 / 582 UE) eine höhere Variabilität innerhalb der Daten zu geben als in der zweiten Studienphase. Ab Text 13 scheinen die Wortstellungsmuster ähnlich oft produziert zu werden (vgl. hierzu die Ausführungen zu Abbildung 41, die die Interaktion der verschiedenen Wortstellungsmuster mit den prozentualen Anteilen zeigt). Eine Ausnahme bilden hier die beiden Texte 22 (51 Wochen / 911 UE) und 23 (60 Wochen / 1051 UE), in denen sehr viele satzwertige Einheiten

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

mit Verbendstellung produziert werden (grüne Linie). TN025 produziert ab dem zweiten Text (5 Wochen / 98 UE) Sätze mit Inversionen. Wie in den Daten für TN024 zeigt sich zu Text 2 einer der höchsten Werte der gesamten Studiendauer für Sätze mit Inversionsstrukturen. TN024 produziert in Text 4 sehr viele Inversionsstrukturen. Der höchste Wert für TN025 findet sich in Text 5 ‚Regeln in Deutschland' (10 Wochen / 205 UE). Hier beginnt TN025 fast alle satzwertigen Einheiten mit lokalen Adverbialangaben wie in den Belegen [36] und [37] im Vorfeld.

[36] *in Deutschland gibt es Viel regeln* (TN025, ‚Regeln in Deutschland')

[37] *in der Bibliothek muss Mann leise sein* (TN025, ‚Regeln in Deutschland')

Während ab Text 2 in jedem Text Sätze mit Inversion produziert werden, realisiert TN025 in einigen Texten keine Sätze, die ausschließlich der Stufe ‚Separation' zugeordnet werden können, etwa in Text 4 (sieben Wochen / 130 UE), 5 (zehn Wochen / 205 UE) und 7 (15 Wochen / 285 UE) oder in Text 10 (23 Wochen / 436 UE). Sätze mit Verbendstellung in Nebensätzen werden, wie bei TN024, ebenfalls ab Text 9 (18 Wochen / 338 UE) produziert. TN025 produziert in Text 12 (28 Wochen / 558 UE) und in Text 21 (48 Wochen / 855 UE) einen integrierten Nebensatz der Profilstufe 5. Erweiterte Partizipialattribute sind in den Daten für TN025 über die gesamte Studiendauer nicht zu finden.

Werden die Rohdaten im zweiten Schritt geglättet betrachtet (vgl. Abbildung 40), ist vor allem die anfängliche Variabilität in der ersten Studienphase deutlich zu sehen. Die Werte für die einzelnen Wortstellungsmuster liegen zum Teil sehr weit auseinander und unterscheiden sich zwischen den einzelnen Zeitpunkten deutlich. Ab etwa Text 10 fällt die angedeutete Verringerung der Variabilität im zeitlichen Kontext der B1-Prüfung (Text 13) noch deutlicher auf. Zwischen Text 12 und Text 17 ist diese am geringsten in der gesamten Studiendauer, um dann gegen Ende der Studie wieder deutlich größer zu werden. Dies hängt vor allem mit einem großen Anstieg der produzierten Sätze mit Verbendstellung und einem Abfallen der Sätze mit Separation zusammen.

Wie für TN024 werden ebenfalls für TN025 die relativen Werte für die Produktion der verschiedenen Wortstellungsmuster analysiert (vgl. Abbildung 41). Die Interaktionen für die Daten der von TN025 produzierten Wortstellungsmuster stellen sich deutlich komplexer als die Daten für TN024 dar. In Abbildung 41 ist zunächst eine durchaus kompetitive Beziehung zwischen den Wortstellungsmustern ‚Finitum' und ‚Inversion' zu erkennen. Dies ist der Fall zwischen Text 1 und Text 12, was der Verteilung in den Daten von TN024 ähnelt. Dazu kommt im Fall von TN025 eine kompetitive Beziehung zwischen den Wortstellungsmustern ‚Separation' und ‚Inversion'. Dies zeigt sich vor allem zwischen den Texten 1 bis 9. Im Anschluss daran entwickeln sich die drei Wortstellungsmuster ‚Separation', ‚Inversion' und ‚Verbendstellung' zunächst sehr ähnlich, während die Werte für ‚Finitum' leicht abnehmen. Nach Text 12 scheinen sich die Variablen eher gemeinsam zu entwickeln bzw. in einen stabileren Zustand einzutreten, der

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

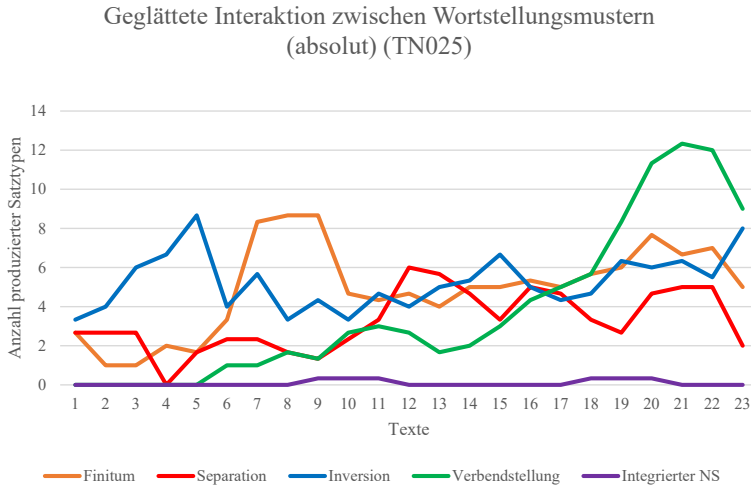


Abbildung 40: Durch Moving Average geglättete Interaktion zwischen den produzierten Wortstellungsmustern für TN025. Der Moving Average hat ein Intervall von 3 und wurde auf der Grundlage der absoluten Anzahl der jeweils produzierten Wortstellungsmuster berechnet.

gleichzeitig mit einem Anstieg der Wortstellungsmuster mit Verbendstellung korreliert. Eine solche Beziehung lässt sich ebenfalls für das Paar ‚Separation' und ‚Verbendstellung' nachzeichnen. Dieser Anstieg von satzwertigen Einheiten der Stufe 4 beginnt zwischen Text 12 und Text 14. Damit passiert dies genau dann, wenn die kompetitive Beziehung zwischen den übrigen Variablen schwächer wird und die Verteilung zwischen den drei übrigen Wortstellungsmustern ausgeglichener erscheint. Sätze mit Verbendstellung werden von TN025 dann vermehrt produziert, wenn das ‚übrige' System aus den Wortstellungsmustern ‚Finitum', ‚Separation' und ‚Inversion' in einen stabileren Zustand eintritt. Die Beziehung der drei übrigen Wortstellungsmuster zum Wortstellungsmuster Verbendstellung lässt sich im Fall von TN025 als konditionell beschreiben.

Insgesamt scheint die Verteilung der einzelnen Wortstellungsmuster vor allem im zweiten Studienteil für TN025 deutlich gleichmäßiger als für TN024 zu sein. Der Anteil an Sätzen mit Verbendstellung steigt kontinuierlich ab Text 5, bis es das meistproduzierte Wortstellungsmuster ist. Gegen Ende der Studie scheint sich das Verhältnis der Wortstellungsmuster zu stabilisieren. Etwa ab Text 19 erreichen die Sätze mit Verbendstellung einen Anteil von knapp 40 % aller Wortstellungsmuster in den Texten.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

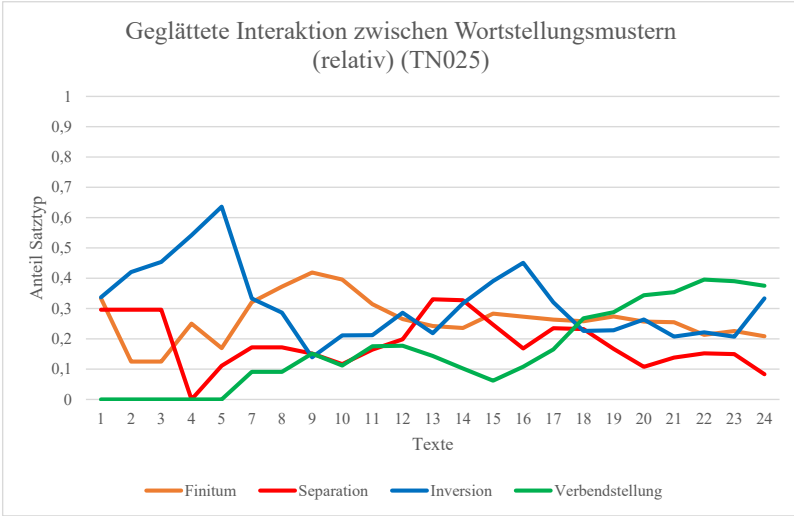


Abbildung 41: Durch Moving Average geglättete Interaktion zwischen den produzierten Wortstellungsmustern für TN025. Der Moving Average hat ein Intervall von 3 und wurde auf der Grundlage der relativen Anzahl der jeweils produzierten Wortstellungsmuster berechnet. Gezeigt werden die prozentuellen Anteile der in den Texten produzierten Wortstellungsmuster.

4.2.4 Grafische Trendanalyse der Interaktionen: Syntaxvariablen

An die grafischen Trend- bzw. Interaktionsanalysen zu den produzierten Wortstellungsmustern anschließend werden nun die Interaktionen der drei untersuchten syntaktischen Variablen in den Blick genommen. Wie zuvor werden die Daten zunächst als Rohwerte und dann mithilfe eines gleitenden Durchschnitts betrachtet. Die Daten sind als *z-scores* normalisiert, da die drei syntaktischen Variablen und die Daten zur erreichten Profilstufe nicht auf derselben Skala vorlagen.[218] Die Variablen ‚Subordinationsrate' und ‚Rate komplexer Nominalphrasen pro MSE' sind im Gegensatz zur Variablen ‚Durchschnittliche Länge einer MSE' verhältnisskaliert und geben prozentuale Anteile wieder. Die Daten zu den erreichten Profilstufen ähneln eher ordinalskalierten Variablen, da es insgesamt nur wenige mögliche Profilstufen und keinerlei genauere Differenzierungen zwischen diesen gibt (zur Diskussion um fehlende Trennschärfe der Profilstufen und der Erwerbsstufen der PT vgl. Kapitel 2.4.3 und ausführlicher

218 Die nicht normalisierten Schaubilder zu den drei analysierten syntaktischen Variablen finden sich in Kapitel 4.2.2 zur grafischen Analyse der Variabilität innerhalb der Daten. Die Rohdaten finden sich in Tabelle 58 und Tabelle 59.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Kapitel 5). Die Daten, die den folgenden Schaubildern zugrunde liegen, wurden in Excel mithilfe der Formel

$$z = \frac{x - \bar{x}}{\sigma}$$

normalisiert. In dieser Gleichung steht x für den tatsächlichen Datenwert und $\bar{x}$ für das arithmetische Mittel (oder den Mittelwert) der Datenreihe. σ steht in der Formel für die Standardabweichung der Datenreihe. Die so ermittelten *z-scores* erlauben es, unterschiedlich skalierte Variablen direkt miteinander zu vergleichen. Ein *z-score*-Datenwert, der über 0 liegt, signalisiert, dass der Wert der betreffenden Variable zu diesem Zeitpunkt über dem Mittelwert liegt. Ein *z-score* unter 0 bedeutet entsprechend einen Ausgangswert unterhalb des Mittelwerts der zugrundeliegenden Datenreihe. Der normalisierte Wert gibt an, um wie viele Standardabweichungen der Wert entfernt vom Mittelwert liegt.

Für die normalisierten Daten zu den syntaktischen Komplexitätsmaßen fällt bei TN024 auf, dass einige Werte sich deutlich von den Mittelwerten der Datenreihe absetzen (vgl. Abbildung 42). Text 4 (nach sieben Wochen / 130 UE) und Text 16 (41 Wochen / 668 UE) weisen im Vergleich zum Mittelwert der Datenreihe deutlich längere satzwertige Einheiten auf als die restlichen Texte. Die Texte 13 (29 Wochen / 550 UE) und 16 fallen deutlich ab, was die erreichte Profilstufe angeht. Vor allem Text 13 ‚B1-Prüfung' fällt auf, da drei von vier Variablen zu diesem Zeitpunkt einen ‚Knick' nach unten aufweisen und unter dem Mittelwert der Datenreihe liegen.

Neben der erreichten Profilstufe (schwarze Linie) liegen in diesem Text kürzere satzwertige Einheiten (rote Linie) und weniger komplexe Nominalphrasen (gelbe Linie) pro MSE im Vergleich zum Mittelwert vor. Einzig die Subordinationsrate (blaue Linie) liegt leicht über dem Mittelwert. Eine weitere Spitze zweier Variablen findet sich in Text 16 ‚Karte' (41 Wochen / 668 UE), der den höchsten Wert für die Variable ‚MLC' und den zweithöchsten Wert für die Rate komplexer Nominalphrasen aufweist. Text 16 fiel bis jetzt nicht als außergewöhnlicher Zeitpunkt auf. Es ging hier darum, anhand eines frei gewählten Phänomens (etwa die weltweite Verbreitung von Fastfood-Ketten) eine entsprechende (geografische) Karte zu entwerfen und diese zu beschreiben. TN024 wählt die Verbindung zwischen Armut und Krieg in der arabischsprachigen Welt. Diese Aufgabenstellung kann den Zusammenhang zwischen längeren satzwertigen Einheiten und komplexen Nominalphrasen erklären, da TN024 an mehreren Stellen Phrasen wie *die arabischen Regionen der Welt* oder *eines der reichsten Arabischen Staaten* (TN024, Text 16 ‚Karte') verwendet, die dann die satzwertigen Einheiten um mehrere Elemente verlängern. Während der Wert für Nebensätze mit Verbendstellung in diesem Text fast bei 0 liegt bzw. etwa dem Mittelwert der Datenreihe entspricht, liegt die erreichte Profilstufe deutlich unter dem Mittelwert der Datenreihe. Der Text wird auf die Profilstufe 1 eingeordnet, was bezogen

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

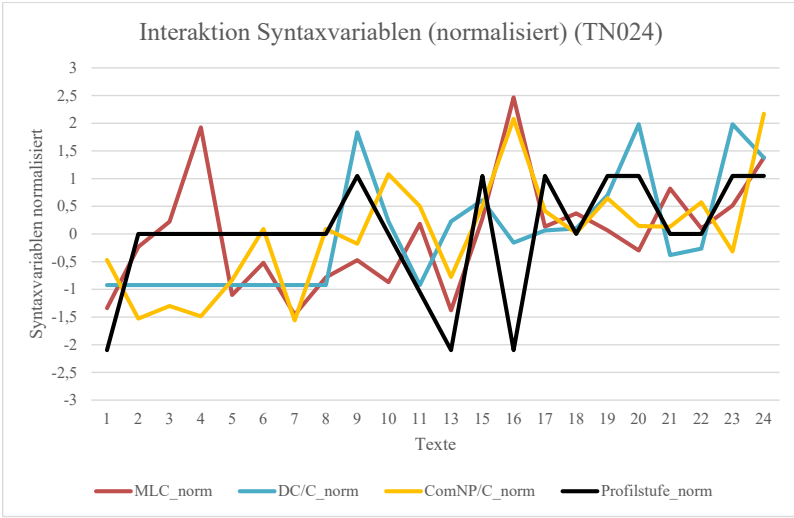


Abbildung 42: Normalisierte Interaktion der Syntaxvariablen für TN024. Gezeigt werden die normalisierten Daten für die durchschnittliche Länge der minimalen satzwertigen Einheiten (MLC_norm), die Subordinationsrate (DC/C_norm) und die komplexen Nominalphrasen pro MSE (ComNP/C_norm). Außerdem wird die durch die Profilanalysen erreichte Profilstufe abgebildet (Profilstufe_norm).

auf die zweitsprachliche Entwicklung von TN024 bis zu diesem Zeitpunkt und im Kontext des Zeitpunkts selbst zumindest überraschend erscheint, da die beiden Komplexitätswerte ‚MLC_norm' und ‚ComNP/C_norm' zum selben Zeitpunkt Spitzenwerte erreichen.

Die geglätteten Daten zeigen zunächst die erwartbare Annäherung an den Mittelwert um 0 (vgl. Abbildung 43). Die Werte der beiden Variablen ‚Subordinationsrate' (blaue Linie) und ‚Rate komplexer Nominalphrasen pro MSE' (gelbe Linie) scheinen sich über die gesamte Studiendauer ähnlich zu entwickeln. Für beide ist in den geglätteten Daten ein leichter Anstieg in den Werten relativ zum Mittelwert zu beobachten. Für die der Werte der Variable ‚MLC_norm_movavg' (bzw. die Länge minimaler satzwertiger Einheiten, rote Linie) gilt dies, mit Ausnahme der Texte zu Beginn der Studie, ebenfalls, wobei hier vor allem Text 4 mit einer Länge von 8,7 Wörtern pro MSE die Werte beeinflusst.

Aufschlussreich ist die Interaktion der Variablen ‚MLC_norm_movavg' und ‚Profilstufe_norm_movavg' (schwarze Linie). Vor allem im ersten Teil der Studie scheinen diese Variablen in einem eher kompetitiven Verhältnis zueinander zu stehen. Diese Interaktion der Variablen ist zwischen Text 4 und Text 10 evident. TN024 schreibt tendenziell eher kürzere, dafür komplexere satzwertige Einheiten oder längere Einheiten, wobei die Gesamttexte dafür auf einer niedrigeren

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

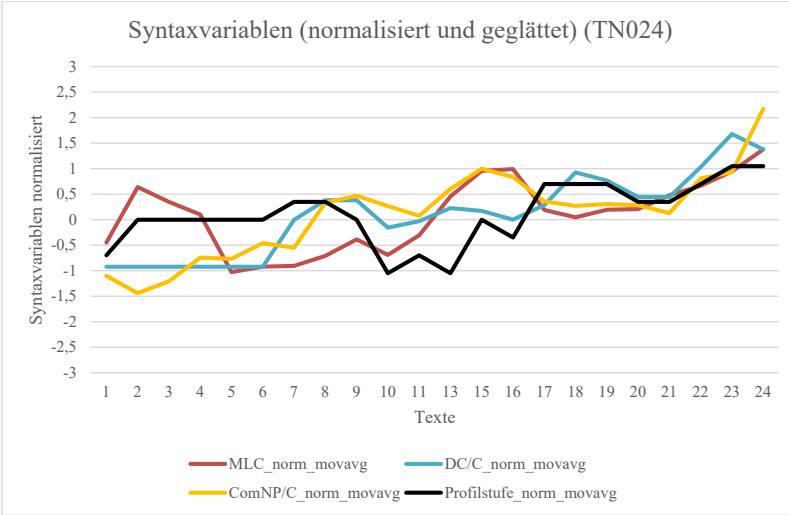


Abbildung 43: Normalisierte und mit einem Moving Average geglättete Interaktion zwischen den drei Syntaxvariablen für TN024. Gezeigt werden die normalisierten Daten für die durchschnittliche Länge der minimalen satzwertigen Einheiten (MLC_norm_movavg), die Subordinationsrate (DC/C_norm_movavg) und die komplexen Nominalphrasen pro MSE (ComNP/C_norm_movavg). Außerdem wird die durch die Profilanalysen ermittelte Profilstufe abgebildet (Profilstufe_norm_movavg). Der Moving Average hat ein Intervall von 3.

Profilstufe eingeordnet werden. Je sicherer TN024 mit komplexeren Strukturen umgehen kann, desto länger werden die satzwertigen Einheiten, die produziert werden. Hier ist zu beachten, dass diese Beobachtung auf geglätteten Daten beruht, die mehrere Datenpunkte gleichzeitig aggregieren. Es handelt sich eher um die Beobachtung eines möglichen Trends als um die Beschreibung einer exakten Beziehung zwischen den Variablen. Im zweiten Teil der Studie ab Zeitpunkt 13 ähneln sich die Kurven der beiden Variablen und ab Text 17 erreicht TN024 ein Deckenniveau mit Profilstufe 4 ‚Verbendstellung'.

Für TN025 fallen innerhalb der normalisierten Daten ebenfalls einige *Peaks* in den Zeitreihendaten der syntaktischen Komplexitätsmaße auf (vgl. Abbildung 44). So erreichen die Werte der Variable ‚MLC_norm' zu den Zeitpunkten 2 (5 Wochen / 98 UE) und 5 (10 Wochen / 205 UE) Höchstwerte, d.h. maximale Abweichungen vom errechneten Mittelwert für diese Variable, die erst im zweiten Studienteil wieder, aber ebenfalls nicht durchgängig, erreicht werden (etwa zu den Zeitpunkten 19 (45 Wochen / 803 UE) und 24 (61 Wochen / 1080 UE). Ein negativer *peak* fällt für ‚MLC_norm' zu Zeitpunkt 10 (23 Wochen / 436 UE) auf, an dem der Wert so deutlich unter dem ermittelten Mittelwert der Variable liegt wie sonst nie im Verlauf der Studie. Anders ausgedrückt: TN025 produziert zu

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

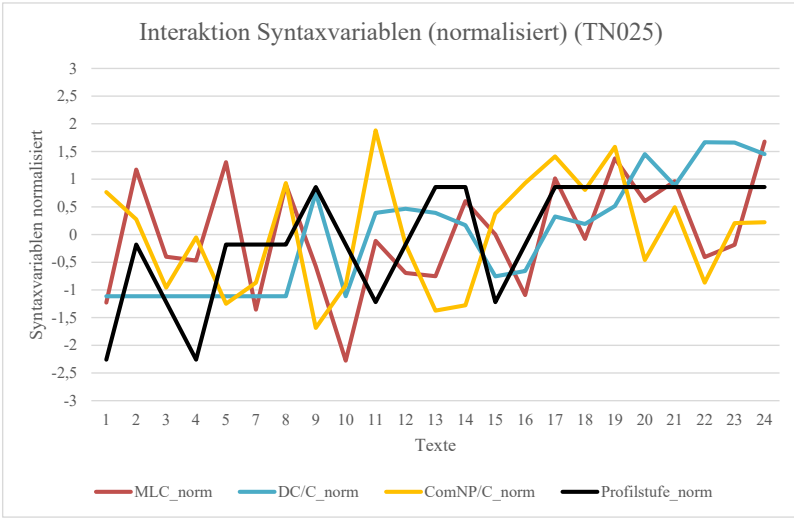


Abbildung 44: Normalisierte Interaktion der Syntaxvariablen für TN025. Gezeigt werden die normalisierten Daten für die durchschnittliche Länge der minimalen satzwertigen Einheiten (MLC_norm), die Subordinationsrate (DC/C_norm) und die komplexen Nominalphrasen pro MSE (ComNP/C_norm). Außerdem wird die durch die Profilanalysen ermittelte Profilstufe abgebildet (Profilstufe_norm).

Zeitpunkt 10 die kürzesten MSE im Studienverlauf. Ein auffallender Höchstwert ergibt sich für die Variable ‚ComNP/C_norm' zu Zeitpunkt 11 (25 Wochen / 486 UE) auf, zu dem der A2-Test geschrieben wird. Die Teilnehmenden sollen hier einen Besuch in Berlin schildern. In diesem Text produziert TN025 zwar nur sechs komplexe Nominalphrasen, da zusätzlich nur wenige MSE realisiert werden, liegt die Rate pro MSE hier im Vergleich zu anderen Datenpunkten aber trotzdem sehr hoch. Hier fallen vor allem Formulierungen wie *mit einem netten teurist* oder *ein schönes Schluss* (beide TN025, Text 11, ‚A2-Test') auf.

In den normalisierten Daten von TN025 fällt auf, dass es zu Zeitpunkt 13 (29 Wochen / 582 UE; B1-Prüfung) nicht zu einem deutlichen Abfall der Variablenwerte wie in den Daten von TN024 kommt. Der Wert der Variable ‚Profilstufe_norm' erreicht einen Höchstwert. Der Text wird der Profilstufe 4 ‚Verbendstellung' zugeordnet. Außerdem kann für die Variablen ‚Profilstufe_norm' und ‚DC/C' (Subordinationsrate) von einer unterstützenden Beziehung über die Studiendauer ausgegangen werden. Die Verlaufskurven der beiden Variablen ähneln sich stark. Dies wird vor allem in Abbildung 45 deutlich. Bis zu Zeitpunkt 15 scheinen die Variablen ‚Profilstufe_norm' und ‚ComNP/C_norm' in einem kompetitiven Verhältnis zueinander zu stehen. Eine Steigerung der Werte der einen Variable korreliert mit einer Verringerung der Werte der anderen Variable. Dies wird vor allem an den zusätzlich geglätteten Daten deutlich (vgl. Abbildung 45).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Abbildung 45 verdeutlicht anhand der normalisierten und mit einem gleitenden Durchschnitt geglätteten Daten, dass die Variablen ‚Profilstufe_norm_movavg' und ‚ComNP/C_norm_movavg' in einem eher kompetitiven Verhältnis zueinanderstehen. Ersichtlich wird dies zwischen Zeitpunkt 5 und Zeitpunkt 9. Die Werte für die erreichte Profilstufe steigen zunächst relativ kontinuierlich, während die Werte für die Rate komplexer Nominalphrasen pro MSE etwas sinken. Ab Zeitpunkt 7 dreht sich diese Entwicklung um. Sehr ähnliche Muster lassen sich außerdem zwischen Zeitpunkt 9 und Zeitpunkt 11 sowie zwischen Zeitpunkt 11 und Zeitpunkt 14 feststellen. Dieser eher kompetitiven Beziehung steht eine unterstützende Beziehung der Variablen ‚Profilstufe_norm_movavg' und ‚DC/C_norm_movavg' gegenüber. Hier laufen die Entwicklungen über die gesamte Studiendauer beinahe parallel.

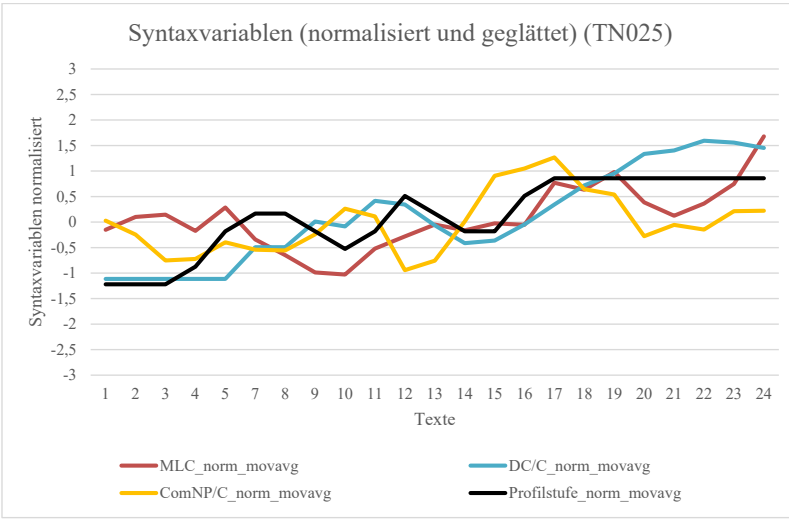


Abbildung 45: Normalisierte und mit einem Moving Average geglättete Interaktion zwischen den drei Syntaxvariablen für TN025. Gezeigt werden die normalisierten Daten für die durchschnittliche Länge der minimalen satzwertigen Einheiten (MLC_norm_movavg), die Subordinationsrate (DC/C_norm_movavg) und die komplexen Nominalphrasen pro MSE (ComNP/C_norm_movavg). Außerdem wird die durch die Profilanalysen ermittelte Profilstufe abgebildet (Profilstufe_norm_movavg). Der Moving Average hat ein Intervall von 3.

Analog zu den Interaktionen der Wortstellungsmuster in den Daten von TN025 (und hier vor allem in Bezug auf den Typ ‚Verbendstellung') lässt sich in den syntaktischen Komplexitätsmaßen für die Variable ‚DC/C_norm_movavg' ein Anstieg der Werte ab Zeitpunkt 14 festhalten, der beinahe bis zum Ende der Studie anhält und erst in den letzten Texten abflacht. Für die Wortstellungsmuster wird eine konditionelle Beziehung zwischen den Wortstellungsmustern

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

‚Finitum', ‚Separation', ‚Inversion' einerseits und den Nebensätzen mit Verbendstellung andererseits angenommen. Dies scheint sich ebenso im kontinuierlichen Anstieg der Werte der Variable ‚DC/C_norm_movavg' (Subordinationsrate) im zweiten Studienteil widerzuspiegeln.

In diesem Abschnitt werden die absoluten Werte und die geglätteten Werte der syntaktischen Komplexitätsmaße für TN024 und TN025 vorgestellt und verglichen. Die Normalisierung der Daten erweist sich bei diesem Vorgehen als sehr nützlich, um vom Mittelwert abweichende *Peaks* in den Daten zu identifizieren, die dann den Blick für Auffälligkeiten in den einzelnen Texten freigeben können. Für beide Teilnehmenden lässt sich ein relativ deutlicher Zusammenhang zwischen der erreichten Profilstufe und der Subordinationsrate feststellen. Je mehr satzwertige Einheiten mit Verbendstellung produziert werden, desto höhere Profilstufen werden erreicht bzw. desto eher wird Profilstufe 4 ‚Verbendstellung' in den Texten erreicht. Die Beziehungen der übrigen Variablen sind etwas komplexer und schwieriger darzustellen. Während sich für TN024 ein negativer Zusammenhang zwischen erreichter Profilstufe und der durchschnittlichen Länge satzwertiger Einheiten aus den Diagrammen herausarbeiten ließ, ergab sich für TN025 ein eher kompetitives Verhältnis zwischen der erreichten Profilstufe und der Rate realisierter komplexer Nominalphrasen pro MSE. Die Beziehung zwischen der Länge satzwertiger Einheiten und der erreichten Profilstufe hingegen scheint sich für TN025 im Gegensatz zu den Daten von TN024 eher durch einen unterstützenden Charakter auszuzeichnen.

4.2.5 Zwischenfazit: Grafische Interpretation

Die grafische Analyse und Interpretation der untersuchten Daten macht einige Punkte sehr deutlich, die so anhand der deskriptiven Statistik noch nicht sichtbar waren. Die durchgeführten grafischen Verfahren legen jeweils spezifische Eigenschaften der lernersprachlichen Daten offen, die sich entweder auf die Variabilität in den Daten oder die Interaktionen der verschiedenen Variablen beziehen. Vor allem die min-max-Graphen können die sonst in den Daten verborgene Variabilität effektiv sichtbar machen. Gerade in longitudinaler Perspektive stellt sich ein solch vielschichtiges Vorgehen als potenziell gewinnbringend dar, da viele zum Teil sehr unterschiedliche Facetten der vorhandenen Daten offengelegt und in Beziehung zueinander gesetzt werden können, die sonst unter Umständen durch die Aggregation der Daten schwieriger zu beobachten wären.

Beim Vergleich der beiden Teilnehmenden TN024 und TN025 fallen unterschiedliche Entwicklungsmuster ins Auge. Dieselben Aufgabenstellungen scheinen bei den beiden Lernenden stark differierende Textrealisierungen zu elizitieren. Dies zeigt sich auf mehreren Ebenen. Über die min-max-Diagramme konnten für beide Teilnehmenden für die untersuchten Variablen jeweils unterschiedliche Phasen größerer Variabilität bzw. Stabilität herausgearbeitet werden. Diese werden in den nächsten Analyseschritten vor allem mit den Ergebnissen der

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Change Point-Analysen kombiniert, um herauszuarbeiten, ob und wie sich Entwicklungssprünge in den Daten einerseits identifizieren und andererseits beschreiben lassen. Es zeigt sich, dass sich für die meisten der untersuchten Variablen die Datenreihen von TN024 durch etwas größere Variabilität auszeichnen. Eine Ausnahme bildet die Datenreihe zur Subordinationsrate, die für TN025 deutlich variabler als für TN024 zu sein scheint. Vor allem die (absoluten) Daten zur Interaktion verschiedener Daten können dann den Blick auf individuelle und differierende Textrealisierungen lenken. Deutlich wird dies etwa beim Text, der von den beiden zur B1-Prüfung geschrieben wird. Während TN024 beinahe ausschließlich Strukturen der Profilstufe 1 ‚Finitum' produziert, obwohl von der teilnehmenden Person davor fast durchgehend mehr Strukturen auf den Profilstufen 3 und 4 als auf Profilstufe 1 produziert werden (vgl. Abbildung 36), nutzt TN025 gerade im Text der B1-Prüfung alle syntaktischen Wortstellungsmuster mit einer fast gleichmäßigen Verteilung (vgl. Abbildung 39). Die Variabilität in den Daten von TN025 ist zu diesem Zeitpunkt insgesamt sehr gering. Der Text von TN024 fällt aus einer longitudinalen Perspektive deutlich aus den bis zu diesem Zeitpunkt beobachteten Entwicklungsmustern heraus. Dies lässt sich unter Umständen mit der Prüfungssituation und unterschiedlichen (impliziten) Strategien der beiden Lernenden erklären. Es ist etwa möglich, dass sich TN024 dafür entscheidet, eher einfachere Sätze zu schreiben, während TN025 sich dafür entscheidet, eher zu zeigen, welche Bandbreite an verschiedenen syntaktischen Strukturen bereits beherrscht wird.

Für beide Teilnehmende ergeben sich für jedes Wortstellungsmuster über die Studiendauer hinweg sehr deutliche Spitzen, d.h. Zeitpunkte, an denen ein Wortstellungsmuster überproportional oft im Vergleich zu den anderen Wortstellungsmuster verwendet wird. Für TN024 fallen hier etwa der angesprochene Zeitpunkt 13, zu dem 13 MSE der Stufe 1 ‚Finitum' realisiert werden, oder Zeitpunkt 20, zu dem zehn MSE der Profilstufe 2 ‚Separation' realisiert werden, auf. TN025 verwendet etwa zu Zeitpunkt 5 15 Strukturen der Profilstufe 2 ‚Separation' und zu Zeitpunkt 10 insgesamt 18 MSE der Profilstufe 1 ‚Finitum', jeweils so viele wie zu keinem anderen Zeitpunkt. Was die Interaktionen zwischen den Variablen angeht, so lässt sich für beide Teilnehmende eine kompetitive Beziehung zwischen den Wortstellungsmustern ‚Finitum' und ‚Separation' ausmachen. Vor allem die relativen Daten zeigen gleichzeitig, dass der Anteil an MSE, die nur als Separation zu klassifizieren sind, über die gesamte Studiendauer für beide Teilnehmenden relativ konstant bleibt. Außerdem finden sich in vielen der Inversionsstrukturen der beiden Lernenden zusätzlich Separationsstrukturen. Dies zeigt sich in Beispielsätzen wie *Danach kaufe ich im Supermarkt ein.* (TN024, Text 2 „Was machen Sie nach dem Deutschkurs?") oder *In den ferien habe Ich viel geschlafen* (TN025, Text 3 „Was hast du gemacht?"). Eine solche Verwendung deutet daraufhin, dass die Lernenden zu sehr frühen Zeitpunkten die eher komplexe Struktur ‚Inversion' weitgehend beherrschen und sie darüber hinaus mit Separationsstrukturen kombinieren können. Während für TN024 zwar ein kontinuierlicher Anstieg des Typs ‚Verbendstellung', aber keine klare

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Beziehung zu den übrigen Typen satzwertiger Einheiten konstruierbar ist, zeigt sich für TN025 zwischen dem Typ ‚Verbendstellung' und den übrigen Typen ‚Finitum', ‚Separation' und ‚Inversion' eine konditionelle Beziehung. Die Werte der Variable ‚Verbendstellung' steigen erst dann, wenn die drei übrigen in einem bestimmten Verhältnis zueinanderstehen.

Verglichen mit den Wortstellungsmustern lassen sich in den Datenreihen der beiden Lernenden zu den übrigen Komplexitätsmaßen (‚MLC', ‚Subordinationsrate', ‚Rate komplexer Nominalphrasen pro MSE') auffällige *Peaks* (hier im Sinne von auffälligen Zeitpunkten in der Gesamtentwicklung) identifizieren, die über die Analyse der Verteilung der unterschiedlichen Typen satzwertiger Einheiten bis dahin nicht auffallen. Zur Identifikation dieser positiven oder negativen Höchstwerte werden zunächst normalisierte (absolute) Daten herangezogen, die Abweichungen vom jeweils errechneten Mittelwert der Datenreihen verdeutlichen. Unterschiede zwischen den Teilnehmenden in Bezug auf die Ausgestaltung einzelner Texte (etwa im Kontext der B1-Prüfung) können durch dieses Vorgehen noch differenzierter sichtbar und an konkreten Daten festgemacht werden. Die normalisierten Daten werden in einem weiteren Schritt durch gleitende Durchschnitte geglättet, wodurch Beziehungen bzw. Interaktionen zwischen den Datenreihen eindeutiger nachvollziehbar werden. Diese stützen für beide Teilnehmende weitgehend die in den vorangegangenen Analyseschritten getroffenen Beobachtungen, konnten aber zusätzliche Unterschiede zwischen beiden Lernenden verdeutlichen.

4.2.6 Generelle und gleitende Korrelationen: Wortstellungsmuster

In diesem Kapitel werden zwei unterschiedliche Korrelationsanalysen zu den untersuchten Variablen als Ergänzung zu den grafischen Analysen der vorangegangenen Kapitel dargestellt. Einerseits sind dies generelle Korrelationen zwischen den verschiedenen Wortstellungsmustern, die die Beziehung über die gesamte Studiendauer verdeutlichen. Und andererseits werden gleitende Korrelationen (*moving correlations*) berechnet, um zu untersuchen, ob sich die Beziehungen zwischen Variablen im Erhebungszeitraum der Studie unter Umständen verändern. Gleitende Korrelationen geben jeweils die Beziehungen zwischen zwei Variablen innerhalb eines bestimmten beweglichen Fensters (*moving window*) wieder und erlauben dadurch, nachzuvollziehen, wie sich die Beziehungen zwischen Variablen im Laufe der Zeit verändern. Sie zeigen, wann die Korrelationen zwischen zwei Variablen stärker werden, abnehmen oder sich keine Korrelationen aufzeigen lassen und die Korrelationswerte gegen null tendieren. Im Rahmen einer komplex-dynamischen Interpretation dieser Beziehungen stehen positive Korrelationen für eher unterstützende Beziehungen (*connected growers*), bei denen sich zwei Variablen nach dem Muster ‚höhere Werte in Variable A entsprechen höheren Werten in Variable B' gemeinsam entwickeln. Negative Korrelationen hingegen deuten aus CDST-Sicht auf

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

eine eher kompetitive Beziehung zwischen den Variablen hin, die sich mit dem Muster ‚höhere Werte in Variable A entsprechen niedrigeren Werten in Variable B / höhere Werte in Variable B entsprechen niedrigeren Werten in Variable A' beschreiben lassen. Im Folgenden werden jeweils die generellen Korrelationen zwischen den Typen satzwertiger Einheiten für die beiden Lernenden und dann die gleitenden Korrelationen berechnet. Im nächsten Kapitel 4.2.7 geht es um die Ergebnisse der Korrelationen zwischen den syntaktischen Komplexitätsmaßen. In den hier vorliegenden Analysen werden für TN024 und TN025 die generellen und gleitenden Korrelationen zwischen folgenden Wortstellungsmustern analysiert:

- ‚Finitum' – ‚Separation'
- ‚Finitum' – ‚Inversion'
- ‚Finitum' – ‚Verbendstellung'
- ‚Separation' – ‚Inversion'
- ‚Separation' – ‚Verbendstellung'
- ‚Inversion' – ‚Verbendstellung'

Die generellen Korrelationen werden mit zusätzlichem Bootstrapping (1000 Bootstraps, Konfidenzintervallniveau 95 %) durchgeführt. D.h., es werden BCa-Konfidenzintervalle auf einem 95 % Konfidenzniveau angegeben (vgl. für die Technik Korrelationen mit zusätzlichem Bootstrap durchzuführen Fußnote 151). Für TN024 ergibt sich aufgrund der 22 in die Analyse eingegangenen Texte folgende Korrelationsmatrix:

Generelle Korrelationen – Wortstellungsmuster (TN024)							
				Finitum	Separation	Inversion	Verbendstellung
Finitum	Korrelation nach Pearson			1	0,067	–0,532*	–0,007
	Signifikanz (2-seitig)				0,769	0,011	0,974
	Bootstrapc	Verzerrung		0	–0,025	0,008	0,018
		Standard Fehler		0	0,220	0,177	0,204
		BCa 95 % Konfidenzintervall	Unterer Wert	.	–0,368	–0,801	–0,413
			Oberer Wert	.	0,458	–0,151	0,488
Separation	Korrelation nach Pearson			0,067	1	–0,055	0,401
	Signifikanz (2-seitig)			0,769		0,810	0,065
	Bootstrapc	Verzerrung		–0,025	0	–0,006	–0,012
		Standard Fehler		0,220	0	0,259	0,201
		BCa 95 % Konfidenzintervall	Unterer Wert	–0,368	.	–0,542	–0,055
			Oberer Wert	0,458	.	0,428	0,758

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Generelle Korrelationen – Wortstellungsmuster (TN024)							
					Finitum	Separation	Inversion	Verbendstellung
Inversion	Korrelation nach Pearson				–0,532*	–0,055	1	0,114
	Signifikanz (2-seitig)				0,011	0,810		0,614
	Bootstrapc	Verzerrung			0,008	–0,006	0	–0,003
		Standard Fehler			0,177	0,259	0	0,255
		BCa 95 % Konfidenzintervall		Unterer Wert	–0,801	–0,542	.	–0,393
				Oberer Wert	–0,151	0,428	.	0,588
Verbendstellung	Korrelation nach Pearson				–0,007	0,401	0,114	1
	Signifikanz (2-seitig)				0,974	0,065	0,614	
	Bootstrapc	Verzerrung			0,018	–0,012	–0,003	0
		Standard Fehler			0,204	0,201	0,255	0
		BCa 95 % Konfidenzintervall		Unterer Wert	–0,413	–0,055	–0,393	.
				Oberer Wert	0,488	0,758	0,588	.
*. Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.								
c. Sofern nicht anders angegeben, beruhen die Bootstrap-Ergebnisse auf 1000 Bootstrap-Stichproben								

Tabelle 60: Generelle Korrelationen – Wortstellungsmuster (TN024)

Die Korrelationsanalysen ergeben mit dem angelegten α-Niveau nur eine einzige signifikante Korrelation zwischen den verschiedenen Wortstellungsmuster (vgl. Tabelle 60):

– Für die Variablen ‚Finitum' und ‚Inversion' wird eine signifikant negative Korrelation mit einem großen Effekt berechnet, $r = -0{,}532$ BCa [–0,801; –0,151], $p = 0{,}011$. Dieser Effekt ist nur knapp als großer Effekt einzustufen.

Eine weitere Korrelation ist fast signifikant, erreicht jedoch nicht das angelegte α-Niveau von 0,05 (vgl. Tabelle 60):

– Die Analyse der Variablen ‚Separation' und ‚Verbendstellung' ergibt keine signifikante Korrelation, aber eine mittlere Effektstärke, $r = 0{,}401$ BCa [–0,055; –0,758], $p = 0{,}065$.

Die übrigen Korrelationsanalysen ergeben keine signifikanten Korrelationen (vgl. Tabelle 60):

– Zwischen den Variablen ‚Finitum' und ‚Separation' ergibt sich keine signifikante Korrelation und zusätzlich nur eine geringe Effektstärke, $r = 0{,}067$ BCa [–0,386; 0,458], $p = 0{,}769$.
– Die Korrelationsanalyse der Variablen ‚Finitum' und ‚Verbendstellung' ergibt keine signifikante Korrelation und nur eine geringe Effektstärke, $r = -0{,}007$ BCa [–0,413; 0,488], $p = 0{,}974$.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

- Die Variablen ‚Separation' und ‚Inversion' korrelieren nicht signifikant miteinander und es ergibt sich nur eine geringe Effektstärke, $r = -0{,}055$ BCa [−0,542; 0,428], $p = 0{,}810$.
- Für die Variablen ‚Inversion' und ‚Verbendstellung' ergibt die Korrelationsanalyse keine signifikante Korrelation und nur eine geringe Effektstärke, $r = 0{,}114$ BCa [−0,393; 0,588], $p = 0{,}614$.

Damit bestätigen die generellen Korrelationsanalysen für TN024 in diesem Schritt vor allem die herausgearbeitete Beziehung der beiden Wortstellungsmuster ‚Finitum' und ‚Inversion' (vgl. Kapitel 4.2.3).

Die Korrelationsmatrix der Wortstellungsmuster zeigt für die Daten von TN025 folgende Ergebnisse:

Generelle Korrelationen – Wortstellungsmuster (TN025)							
				Finitum	Separation	Inversion	Verbendstellung
Finitum	Korrelation nach Pearson			1	−0,081	−0,068	0,250
	Signifikanz (2-seitig)				0,713	0,758	0,250
	Bootstrapc	Verzerrung		0	0,028	−0,012	0,054
		Standard Fehler		0	0,256	0,264	0,258
		95 % Konfidenzintervall	Unterer Wert	1	−0,483	−0,584	−0,125
			Oberer Wert	1	0,565	0,415	0,796
Separation	Korrelation nach Pearson			−0,081	1	−0,293	0,367
	Signifikanz (2-seitig)			0,713		0,175	0,085
	Bootstrapc	Verzerrung		0,028	0	0,018	−0,008
		Standard Fehler		0,256	0	0,197	0,205
		95 % Konfidenzintervall	Unterer Wert	−0,483	1	−0,629	−0,075
			Oberer Wert	0,565	1	0,156	0,739
Inversion	Korrelation nach Pearson			−0,068	−0,293	1	0,035
	Signifikanz (2-seitig)			0,758	0,175		0,875
	Bootstrapc	Verzerrung		−0,012	0,018	0	0,021
		Standard Fehler		0,264	0,197	0	0,219
		95 % Konfidenzintervall	Unterer Wert	−0,584	−0,629	1	−0,329
			Oberer Wert	0,415	0,156	1	0,520
Verbendstellung	Korrelation nach Pearson			0,250	0,367	0,035	1
	Signifikanz (2-seitig)			0,250	0,085	0,875	
	Bootstrapc	Verzerrung		0,054	−0,008	0,021	0
		Standard Fehler		0,258	0,205	0,219	0
		95 % Konfidenzintervall	Unterer Wert	−0,125	−0,075	−0,329	1
			Oberer Wert	,0796	,0739	0,520	1
c. Sofern nicht anders angegeben, beruhen die Bootstrap-Ergebnisse auf 1000 Bootstrap-Stichproben							

Tabelle 61: Generelle Korrelationen – Wortstellungsmuster (TN025)

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Die Korrelationsanalysen der Wortstellungsmuster für die 23 Texte von TN025 ergeben keine signifikanten Korrelationen und nur geringe bis mittlere Effektstärken (vgl. Tabelle 61):

- Die Korrelationsanalyse der Variablen ‚Finitum' und ‚Separation' ergibt keine signifikante Korrelation mit nur einer geringen Effektstärke, $r = -0{,}081$ BCa [−0.483; 0,565], $p = 0{,}713$.
- Für die Variablen ‚Finitum' und ‚Inversion' ergibt sich keine signifikante Korrelation und nur eine geringe Effektstärke, $r = -0{,}068$ BCa [−0,584; 0,415], $p = 0{,}758$.
- Die Korrelationsanalyse der Variablen ‚Finitum' und ‚Verbendstellung' ergibt keine signifikante Korrelation und eine mittlere Effektstärke, $r = 0{,}250$ BCa [−0,125; 0,796], $p = 0{,}250$.
- Für die Variablen ‚Separation' und ‚Inversion' kann keine signifikante Korrelation und eine mittlere Effektstärke festgestellt werden, $r = -0{,}293$ BCa [−0,629; 0,156], $p = 0{,}175$.
- Die Analyse der Variablen ‚Separation' und ‚Verbendstellung' ergibt ebenfalls keine signifikante Korrelation und eine mittlere Effektstärke, $r = 0{,}367$ BCa [−0,075; 0,739], $p = 0{,}085$. In diesem Fall tendiert die Korrelation entgegen des geforderten α-Niveaus, erreicht es jedoch nicht ganz.
- Und zuletzt ergibt die Analyse der Variablen ‚Inversion' und ‚Verbendstellung' keine signifikante Korrelation und nur eine geringe Effektstärke, $r = 0{,}035$ BCa [−0,329; 0,520], $p = 0{,}875$.

Für beide Teilnehmenden erreichen die Variablen ‚Separation' und ‚Verbendstellung' beinahe eine signifikante (positive) Korrelation und in beiden Fällen ergeben sich mittlere Effektstärken. Dies deutet auf einen (möglichen) Zusammenhang hin, der in den bisher durchgeführten Analysen in dieser Form nicht sichtbar wird. Die stark kompetitive Beziehung zwischen den Variablen ‚Finitum' und ‚Inversion' zeigt sich nur in den Daten von TN024.

Eine wichtige Eigenschaft longitudinaler Daten ist die Veränderlichkeit der gegenseitigen Beziehungen von Variablen im Zeitverlauf. Falls sich für bestimmte Variablenpaare in Korrelationsanalysen keine signifikanten Korrelationen ergeben, könnte dies unter anderem daran liegen, dass in einer Phase eine stark positive und in einer anderen Phase eine stark negative Beziehung der Variablen zueinander vorliegt, was dann dazu führen könnte, dass sich diese Ergebnisse bei der Berechnung einer Korrelation über die gesamte Studiendauer aufheben. Um die Beziehungen kleinteiliger im zeitlichen Verlauf analysieren zu können, werden nun die Ergebnisse der gleitenden Korrelationen vorgestellt. Für jede der gleitenden Korrelationen wurde ein gleitendes Fenster von fünf Messpunkten gewählt (für eine Diskussion um die angemessene Größe von gleitenden Fenstern vgl. Kapitel 3.3.4). Für jedes dieser Fenster wird ein Korrelationskoeffizient berechnet. Im Gegensatz zu generellen Korrelationen wird für die folgenden gleitenden Korrelationen darauf verzichtet, anzugeben, ob es sich bei der für dieses Fenster berechneten Korrelation um eine statistisch signifikante

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

handelt. Dies hat vor allem damit zu tun, dass es bei gleitenden Korrelationen darum geht, Trends in den Daten nachzuvollziehen. Hier ist daran zu erinnern, dass im Falle von Korrelationen der Korrelationseffizient Pearson's *r* gleichzeitig die Effektstärke der Korrelation angibt. Die Effektstärken können durchaus Hinweise darauf geben, wie stark die Beziehungen zwischen den Datenreihen in den jeweiligen Fenstern tatsächlich sind. Wie bei generellen Korrelationen können die Korrelationskoeffizienten einer gleitenden Korrelation Werte zwischen –1 und 1 erreichen, wobei die Beziehung zwischen den Variablen stärker ist, je mehr sich der Wert –1 oder 1 nähert. Ein Wert, der Richtung 0 tendiert, signalisiert dementsprechend, dass es keine quantifizierbare Beziehung zwischen den Variablen gibt. Gerade für die Wortstellungsmuster kann die Verwendung bestimmter Satztypen persönliche Schreibpräferenzen illustrieren. Daher werden im Folgenden die sechs Korrelationen von TN024 und TN025 jeweils nebeneinander abgebildet und dann vergleichend besprochen. Da bei diesen, wie bei den grafischen Trendanalysen, *moving windows* zum Einsatz kommen und daher Datenpunkte aggregiert werden, wird auf die Angabe von exakten Zeitangaben sowie auf die Angabe der besuchten Unterrichtseinheiten verzichtet.

Gleitende Korrelation: ‚Finitum' – ‚Separation'

Abbildung 46 zeigt die gleitende Korrelation der beiden Wortstellungsmuster ‚Finitum' und ‚Separation' für TN024 (oberes Diagramm) und TN025 (unteres Diagramm). In Abbildung 46 wird deutlich, dass die beiden Variablen über die

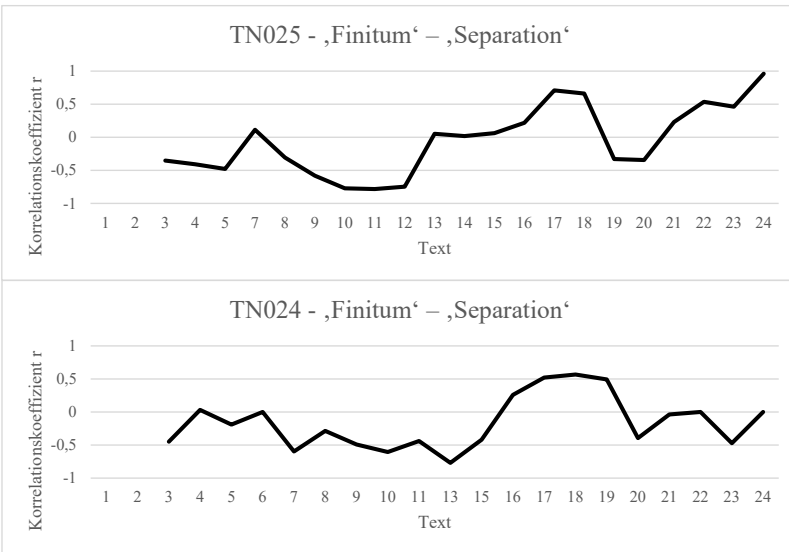


Abbildung 46: Gleitende Korrelation ‚Finitum' – ‚Separation' für TN024 (oben) und TN025 (unten)

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

gesamte Studiendauer, aber vor allem im ersten Studienteil, eher negativ korrelieren bzw. gegen 0 tendieren. Dies deutet darauf hin, dass keine beobachtbare Beziehung zwischen den beiden Wortstellungsmustern vorliegt. Im zweiten Studienteil fällt zunächst eine Phase gemeinsamer Entwicklung der beiden Wortstellungsmuster auf, die sich sowohl für TN024 als auch für TN025 aufzeigen lässt. Für TN024 beginnt diese Phase zu Zeitpunkt 13 und für TN025 etwas früher, nämlich zu Zeitpunkt 11. In dieser Phase korrelieren die beiden Wortstellungsmuster immer stärker, bis die Kurven der Korrelationen ab Zeitpunkt 18 für TN024 und ab Zeitpunkt 17 für TN025 wieder etwas abfallen. Für TN025 fällt zusätzlich eine Phase sehr starker Korrelation ab Zeitpunkt 22 bis zum Ende der Studie auf, in der durchgehend große Effektstärken zu beobachten sind.

Die generelle bzw. die auf die gesamte Studiendauer bezogene Korrelation ist für das Variablenpaar ‚Finitum' und ‚Separation' für beide Lernenden nicht signifikant. Vor allem anhand des Diagramms für TN025 wird deutlich, dass dies an den alternierenden Phasen zwischen negativen und positiven Korrelationen festgemacht werden könnte.

Gleitende Korrelation: ‚Finitum' – ‚Inversion'

Die nächste Abbildung visualisiert für TN024 und TN025 die gleitenden Korrelationen zwischen den Variablen ‚Finitum' und ‚Inversion' (vgl. Abbildung 47). Für die gleitenden Korrelationen zwischen den Wortstellungsmustern ‚Finitum' und ‚Inversion'unterscheiden sich die Diagramme der beiden Lernenden deutlicher. Die Werte für TN024 bleiben fast über die gesamte Studiendauer im negativen Bereich. Es ist insgesamt von einer negativen Korrelation der beiden Variablen auszugehen. Für das Diagramm von TN025 stellt sich ein deutlich davon abweichendes Bild dar. Die Beziehung der Variablen liegt zwar immer wieder im negativen Bereich. Insgesamt zeigen sich lange Phasen, in denen die Variablen positiv miteinander korreliert sind, z. B. zwischen Zeitpunkt 8 (17 Wochen / 300 UE) und Zeitpunkt 13 (29 Wochen / 582 UE) und zwischen Zeitpunkt 18 (45 Wochen / 793 UE) und Zeitpunkt 21 (48 Wochen / 855 UE). Damit bestätigen die gleitenden Korrelationen die bisher gemachten Beobachtungen weitgehend.

Die generelle Korrelation zwischen den Variablen ‚Finitum' und ‚Inversion' ist die einzige Korrelation, die für TN024 als signifikant eingestuft wird (vgl. Tabelle 60). Für TN025 erreicht die generelle Korrelation nicht das signifikante α-Niveau (vgl. Tabelle 61). Während für TN024 relativ deutlich ist, dass die Werte fast dauerhaft negativ sind, fällt für TN025 auf, dass die Werte der gleitenden Korrelation zweimal sehr deutlich positiv und in der übrigen Zeit negativ sind und sich deshalb unter Umständen keine generelle Korrelation ergibt.

Gleitende Korrelation: ‚Finitum' – ‚Verbendstellung'

Wie bei der vorangegangenen Analyse zeigen sich in den beiden Diagrammen für die gleitenden Korrelationen der Wortstellungsmuster ‚Finitum' und ‚Verbendstellung' für die beiden Teilnehmenden stark voneinander abweichende Entwicklungsmuster (vgl. Abbildung 48).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

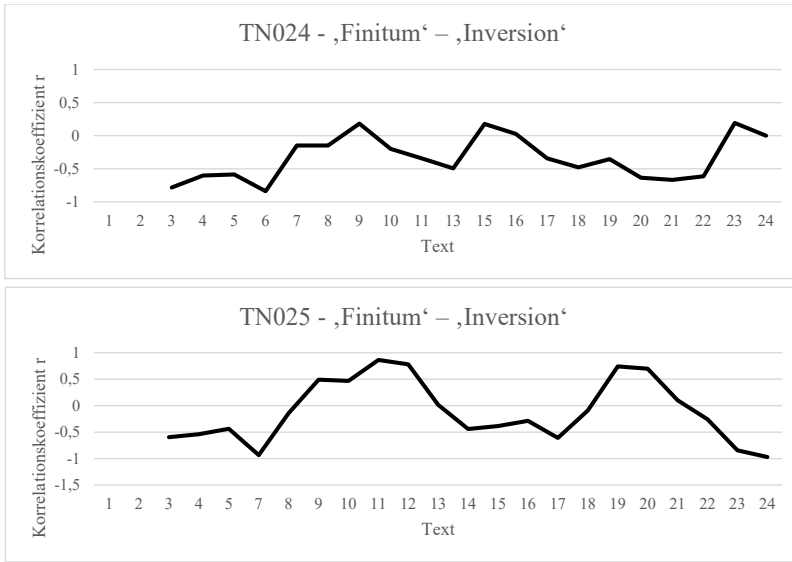


Abbildung 47: Gleitende Korrelation ‚Finitum' – ‚Inversion' für TN024 (oben) und TN025 (unten)

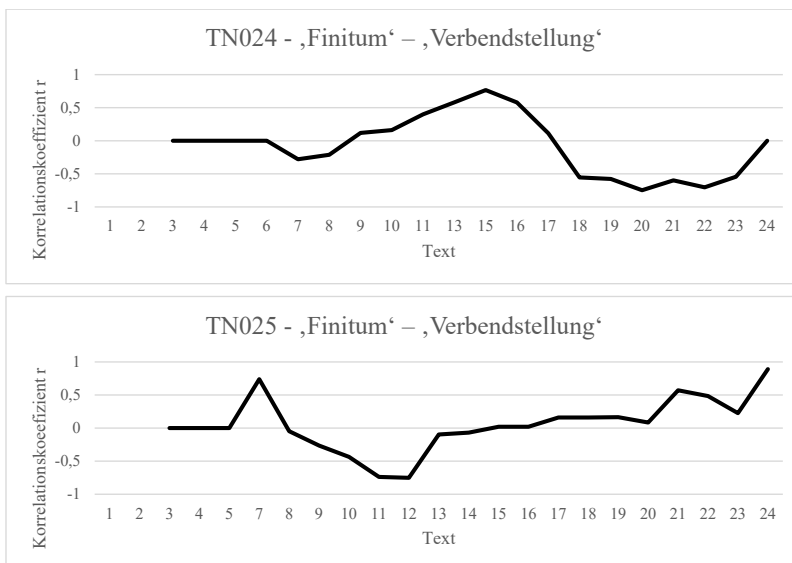


Abbildung 48: Gleitende Korrelation ‚Finitum' – ‚Verbendstellung' für TN024 (oben) und TN025 (unten)

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Zunächst zeigt sich für beide eine Phase, die bei exakt 0 liegt, da noch keine satzwertigen Einheiten mit Verben in Endstellung produziert werden. Für TN024 beginnt ab Zeitpunkt 7 (15 Wochen / 292 UE) eine Phase, in der die Korrelation zwischen den Variablen kontinuierlich größer wird und bis Zeitpunkt 13 (29 Wochen 550 UE) anwächst. In dieser Phase lässt sich eine positive Korrelation beobachten. Je mehr satzwertige Einheiten auf Stufe 1 ‚Finitum' realisiert werden, desto mehr MSE mit Verben in Endstellung finden sich. Ab Zeitpunkt 13 wird diese Korrelation negativ, d.h., die beiden Variablen stehen in einem eher kompetitiven Verhältnis. Je mehr MSE auf Profilstufe 1 produziert werden, desto weniger MSE werden auf Profilstufe 4 produziert und andersherum. Für TN025 liegt der Wert der gleitenden Korrelation fast über die gesamte Studiendauer beinahe bei null. Die beiden Variablen scheinen über zahlreiche Phasen nicht miteinander korreliert zu sein. Auffällig ist der kurze *Peak* zu Zeitpunkt 6, dem eine Phase eher negativer Korrelation von Zeitpunkt 7 bis zu Zeitpunkt 12 folgt, bis sich die Werte wieder null annähern.

Obwohl die Beziehung der Variablen ‚Finitum' und ‚Verbendstellung' für TN024 und TN205 als statistisch nicht signifikant eingestuft wird (vgl. Tabellen 60 und 61), ergeben sich relevante Muster, die sich zusätzlich zwischen beiden Lernenden deutlich unterscheiden.

Gleitende Korrelation: ‚Separation' – ‚Inversion'

In Abbildung 49 sind die gleitenden Korrelationen für die Wortstellungsmuster ‚Separation' und ‚Inversion' zu sehen.

Für TN024 ist die gleitende Korrelation über weite Teile der Studiendauer positiv und zwar ab etwa Zeitpunkt 6 (12 Wochen / 231 UE) bis zu Zeitpunkt 20 (45 Wochen / 771 UE). Dies passt zu den bisher herausgearbeiteten Interaktionen der Wortstellungsmuster (vgl. Abbildung 37). Dann fällt jedoch auf, dass die Korrelation sich sehr schnell in den negativen Bereich verschiebt und dort hohe Effektstärken erreicht. Dies ist ab Zeitpunkt 20 der Fall und zieht sich bis zum Ende der Studie (Zeitpunkt 24, 61 Wochen / 1037 UE). Die gleitende Korrelation für TN025 bleibt über die gesamte Studiendauer im negativen Bereich und zeichnet sich vor allem ab Zeitpunkt 10 durch größere Effektstärken aus. Die Beziehung der beiden Variablen besitzt einen stark kompetitiven Charakter, der bis zu diesem Punkt nicht deutlich herausgearbeitet werden konnte (vgl. die Beschreibung zu Abbildung 40). Während die bisher gezeigten kompetitiven Beziehungen sich durch einen Wechsel der Korrelationsrichtung zwischen positiven und negativen Korrelationen ausgezeichnet haben, scheint in dieser Korrelation durchgehend ein Wortstellungsmuster zu Ungunsten des anderen produziert zu werden. In den grafischen Trendanalysen wird relativ deutlich, dass der Anteil an MSE mit ausschließlich Separationskonstruktionen für TN025 über die gesamte Studiendauer im Vergleich zu Sätzen mit Inversionsstrukturen konstant auf einem (im Vergleich niedrigen) Niveau bleibt (vgl. Abbildungen 38 und 41). Die Datenreihen der beiden Lernenden unterscheiden sich in diesen

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

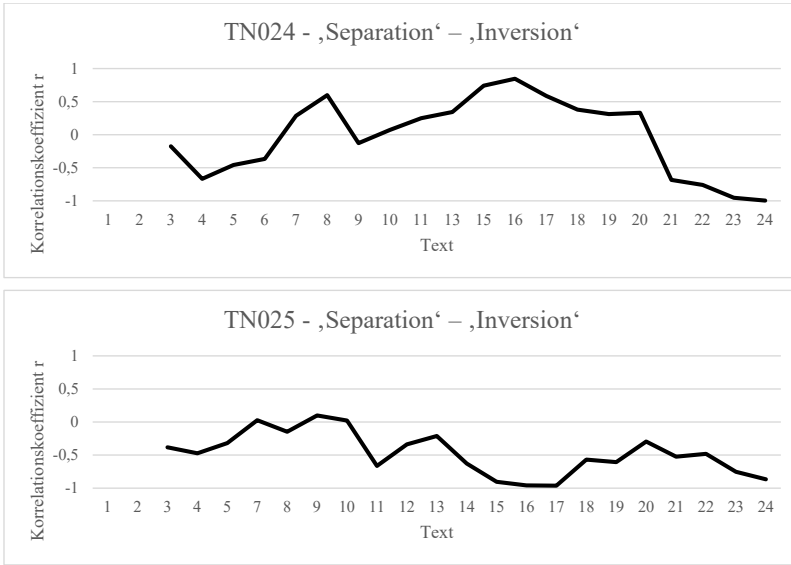


Abbildung 49: Gleitende Korrelation ‚Separation' – ‚Inversion' für TN024 (oben) und TN025 (unten)

Korrelationen sehr deutlich und lassen auf grundsätzlich verschiedene Schreibpräferenzen bei der Auswahl dieser beiden Wortstellungsmuster schließen.

Die generelle Korrelationsanalyse ergibt für die beiden Variablen ‚Separation' und ‚Inversion' für beide Lernende eine nicht signifikante Beziehung (vgl. Tabelle 60 und 61), was zumindest hinsichtlich der Daten für TN025 überraschend erscheint, da sich die gleitenden Korrelationen durchgehend im negativen Bereich finden.

Gleitende Korrelation: ‚Separation' – ‚Verbendstellung'

Die gleitenden Korrelationen zwischen den Wortstellungsmustern ‚Separation' und ‚Verbendstellung' werden in Abbildung 50 gezeigt. In den gleitenden Korrelationen lässt sich eine Phase festmachen, in der es zu keiner sichtbaren Beziehung zwischen den beiden Wortstellungsmustern kommt und die Korrelationen gegen null tendieren. Dies ist etwa bis zu den Zeitpunkten 8 bzw. 9 der Fall, was wiederum darauf zurückzuführen ist, dass die ersten MSE mit Verbendstellung erst ab Zeitpunkt 9 produziert werden. Während die Korrelation zwischen ‚Separation' und ‚Verbendstellung' für TN024 ab diesem Zeitpunkt zunächst negativ wird, steigt die Korrelation für TN025 ab diesem Zeitpunkt ins Positive und bleibt dies mit wenigen Ausschlägen nach unten bis zum Ende der Studiendauer. Grundsätzlich kann festgehalten werden, dass die gleitenden Korrelationen der beiden Lernenden sich in diesem Fall stark ähneln. Spätestens mit

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Beginn der zweiten Studienphase ab Zeitpunkt 13 setzt eine Phase ein, in der die Korrelationen jeweils durch eine positive Beziehung bestimmt sind, in der kontinuierlich größere Effektstärken beobachtbar sind.

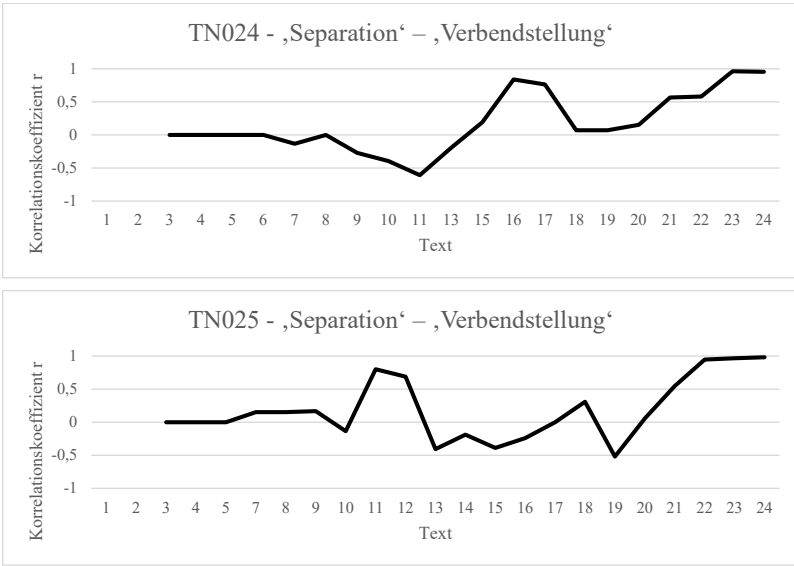


Abbildung 50: Gleitende Korrelation ‚Separation' – ‚Verbendstellung' für TN024 (oben) und TN025 (unten)

Für die beiden Variablen ‚Separation' und ‚Verbendstellung' sind die generellen Korrelationen für beide Lernenden nicht signifikant, wobei sie für TN024 nur knapp das geforderte α-Niveau verfehlt.

Gleitende Korrelation: ‚Inversion' – ‚Verbendstellung'

Die letzte der gleitenden Korrelationen, die in Bezug auf die Wortstellungsmuster vorgestellt wird, ist die Korrelation zwischen ‚Inversion' und ‚Verbendstellung' (vgl. Abbildung 51). Wie in den übrigen Korrelationen, die die Variable ‚Verbendstellung' beinhalten, zeichnen sich diese Korrelationen durch eine anfängliche Phase aus, in der keine Beziehung festgestellt werden kann. Nach dieser Phase unterscheiden sich die berechneten gleitenden Korrelationen für TN024 und TN025 deutlich. Die Daten für TN024 zeigen ab Zeitpunkt 7 eine gleitende Korrelation, die zunächst etwas schwächer wird und bis zu Zeitpunkt 11 gegen null tendiert, um ab diesem Zeitpunkt bis zu Zeitpunkt 18 kontinuierlich stärker zu werden. Zum Ende der Studie fällt die Beziehung wieder stark ab, bis sich eine negative Korrelation mit großer Effekstärke zeigt (vgl. vor allem Zeitpunkt 23 und 24). Die Entwicklung in den Daten von TN025 zeichnet sich durch

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

eine größere Variabilität aus. Die Korrelation steigt immer wieder stark an, etwa zwischen Zeitpunkten 10 und 13 oder 18 und 20. Sie fällt jedoch zugleich immer wieder deutlich ab, etwa zwischen Zeitpunkt 13 und Zeitpunkt 18, in der die Korrelation negativ wird und die Effektgröße immer weiter zunimmt.

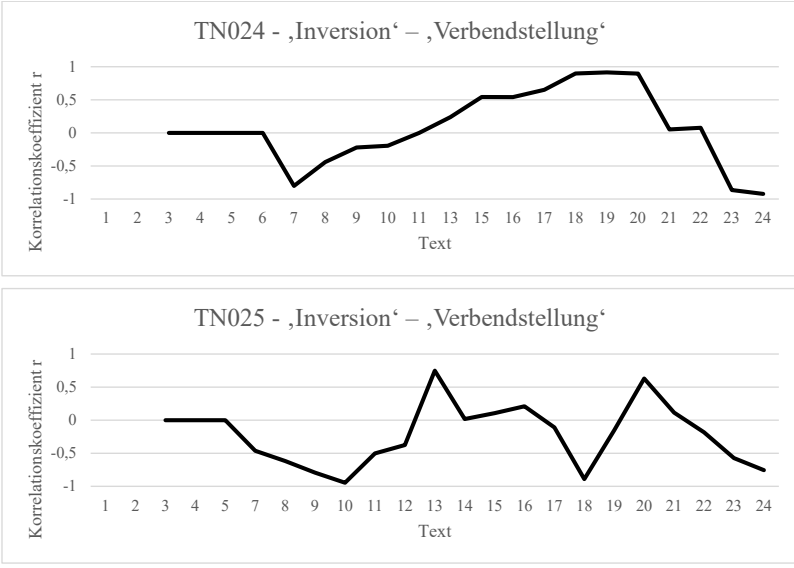


Abbildung 51: Gleitende Korrelation ‚Inversion' – ‚Verbendstellung' für TN024 (oben) und TN025 (unten)

Die generelle Korrelation der Variablen ‚Inversion' und ‚Verbendstellung' ergibt keine signifikante Korrelation und nur eine geringe Effektstärke. Vor allem für TN025 wird deutlich, dass die unterschiedlichen Phasen der sich ändernden Beziehungen hierfür ausschlaggebend sein könnten.

4.2.7 Generelle und gleitende Korrelationen: Syntaxvariablen

Im Folgenden werden für TN024 und TN025 folgende generelle und gleitende Korrelationen anhand der normalisierten Daten (vgl. Kapitel 4.2.4) vorgestellt:

- ‚Durchschnittliche Länge einer minimalen satzwertigen Einheit' (MLC) – ‚Subordinationsrate'
- ‚MLC' – ‚Rate komplexer NPs pro MSE'
- ‚Subordinationsrate' – ‚Rate komplexer NPs pro MSE'

Wie bei den Korrelationen der Wortstellungsmuster werden die generellen Korrelationen zu den syntaktischen Variablen mit zusätzlichem Bootstrapping

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

(1000 Bootstraps, 95 % Konfidenzintervallniveau) durchgeführt. Für TN024 ergibt dies auf Grundlage der 22 eingeschlossenen Texte folgende Korrelationsmatrix:

Generelle Korrelationen – Syntaxvariablen (TN024)						
				MLC_norm	DC/ C_norm	ComNP/ C_norm
MLC_ norm	Korrelation nach Pearson			1	0,152	0,433*
	Signifikanz (2-seitig)				0,500	0,044
	Bootstrapc	Verzerrung		0	0,010	–0,025
		Standard Fehler		0	0,187	0,279
		95 % Konfi- denzintervall	Unterer Wert	1	–0,193	–0,219
			Oberer Wert	1	0,555	0,855
DC/C_ norm	Korrelation nach Pearson			0,152	1	0,393
	Signifikanz (2-seitig)			0,500		0,071
	Bootstrapc	Verzerrung		0,010	0	–0,001
		Standard Fehler		0,187	0	0,169
		95 % Kon- fidenzintervall	Unterer Wert	–0,193	1	0,039
			Oberer Wert	0,555	1	0,704
ComNP/ C_norm	Korrelation nach Pearson			0,433*	0,393	1
	Signifikanz (2-seitig)			0,044	0,071	
	Bootstrapc	Verzerrung		–0,025	–0,001	0
		Standard Fehler		0,279	0,169	0
		95 % Kon- fidenzintervall	Unterer Wert	–0,219	0,039	1
			Oberer Wert	0,855	0,704	1
*. Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.						
c. Sofern nicht anders angegeben, beruhen die Bootstrap-Ergebnisse auf 1000 Bootstrap-Stichproben						

Tabelle 62: Generelle Korrelationen – Syntaxvariablen (TN024)

In den Korrelationsanalysen der normalisierten syntaktischen Komplexitätsmaße ‚Durchschnittliche Länge minimaler satzwertiger Einheiten' (MLC_ norm), ‚Subordinationsrate' (DC/C_norm) und ‚Rate komplexer NPs pro MSE' (ComNP/C_norm) für TN024 ergibt sich eine signifikante Korrelation:

– Die Analyse der Variablen ‚MLC_norm' und ‚ComNP/C_norm' ergibt eine signifikant positive Korrelation mit mittlerer Effektstärke, r = 0,433 BCa [–0,219; 0,855], p = 0,044.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Eine weitere Korrelation ist beinahe signifikant und zeigt eine mittlere Effektstärke, erreicht aber das geforderte α-Niveau nicht ganz:

– Für die Variablen ‚DC/C_norm' und ‚ComNP/C_norm' ergibt sich keine signifikante Korrelation, aber eine mittlere Effektstärke, $r = 0{,}393$ BCa [0,039; 0,704], $p = 0{,}071$.

Die dritte Korrelation ist nicht signifikant:

– Die Korrelationsanalyse der Variablen ‚MLC_norm' und ‚DC/C_norm' ergibt keine signifikante Korrelation und eine geringe Effektstärke, $r = 0{,}152$ BCa [–0,193; 0,555], $p = 0{,}500$.

Die signifikant positive Korrelation zwischen Länge einer durchschnittlichen MSE und der Rate komplexer NPs pro MSE gibt einen Hinweis darauf, dass TN024 mehr komplexe NPs realisiert, wenn längere MSE geschrieben werden.

Für TN025 ergibt sich für die durchgeführten Korrelationen anhand der 23 eingeschlossenen Texte folgende Korrelationsmatrix:

Generelle Korrelationen – Syntaxvariablen (TN025)						
				MLC_norm	DC/ C_norm	ComNP C_norm
MLC_ norm	Korrelation nach Pearson			1	0,285	0,296
	Signifikanz (2-seitig)				0,188	0,170
	Bootstrapc	Verzerrung		0	–0,018	–0,007
		Standard Fehler		0	0,199	0,180
		95 % Konfidenzintervall	Unterer Wert	1	–0,176	–0,077
			Oberer Wert	1	0,631	0,618
DC/C_ norm	Korrelation nach Pearson			0,285	1	0,020
	Signifikanz (2-seitig)			0,188		0,928
	Bootstrapc	Verzerrung		–0,018	0	0,004
		Standard Fehler		0,199	0	0,165
		95 % Konfidenzintervall	Unterer Wert	–0,176	1	–0,309
			Oberer Wert	0,631	1	0,352
ComNP/ C_norm	Korrelation nach Pearson			0,296	0,020	1
	Signifikanz (2-seitig)			0,170	0,928	
	Bootstrapc	Verzerrung		–0,007	0,004	0
		Standard Fehler		0,180	0,165	0
		95 % Konfidenzintervall	Unterer Wert	–0,077	–0,309	1
			Oberer Wert	0,618	0,352	1
c. Sofern nicht anders angegeben, beruhen die Bootstrap-Ergebnisse auf 1000 Bootstrap-Stichproben						

Tabelle 63: Generelle Korrelationen – Syntaxvariablen (TN025)

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Für die Daten von TN025 ergeben die Korrelationsanalysen keine signifikanten Korrelationen:

- Für die Variablen ‚MLC_norm' und ‚DC/C_norm' ergibt sich keine signifikante Korrelation und eine mittlere Effektgröße, r = 0,285 BCa [−0,176; 0,631], p = 0,188.
- Die Analyse der Variablen ‚MLC_norm' und ‚ComNP/C_norm' zeigt keine signifikante Korrelation und eine mittlere Effektgröße, r = 0,296 BCa [−0,077; 0,618], p = 0,170.
- Für die Variablen ‚DC/C_norm' und ‚ComNP/C_norm' ergibt sich keine signifikante Korrelation und eine geringe Effektstärke, r = 0,020 BCa [−0,309; 0,352], p = 0,928.

Der Zusammenhang, der am ehesten überindividuell zwischen beiden Lernenden hinsichtlich der Korrelationsanalysen der Syntaxvariablen besteht, ist die positive Beziehung zwischen den Variablen ‚Durchschnittliche Länge einer MSE' und ‚Rate komplexer NPs pro MSE'. Für TN024 ergibt sich hier eine signifikante Korrelation mit mittlerer Effektgröße und für TN025 zumindest eine mittlere Effektgröße. Eine größere Anzahl komplexer Nominalphrasen hängt demnach mit längeren minimalen satzwertigen Einheiten zusammen.

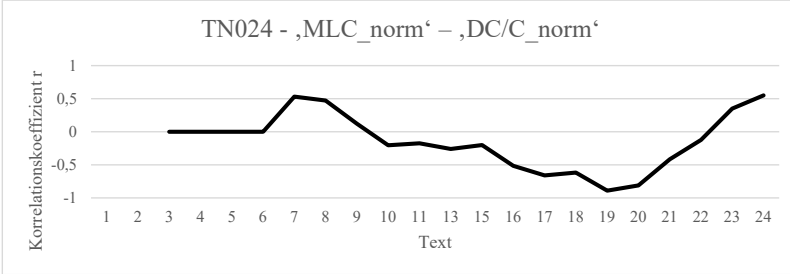
Wie für die verschiedenen Wortstellungsmuster sollen nun die gleitenden Korrelationen die sich im Zeitverlauf ändernden Beziehungen zwischen den verwendeten Komplexitätsmaßen aufzeigen. Dabei liegen allen gleitenden Korrelationen die normalisierten Daten zugrunde (vgl. Kapitel 4.2.4 zur Normalisierung der Daten).

‚Durchschnittliche Länge einer MSE' (MLC) – ‚Subordinationsrate'

Die erste der gleitenden Korrelationen, die in den Blick genommen wird, ist die Korrelation zwischen der durchschnittlichen Länge minimaler satzwertiger Einheiten und der Subordinationsrate (vgl. Abbildung 52). Wie in den Abbildungen, in denen die Variable ‚Verbendstellung' eine Rolle spielte, zeigen sich in den Daten zur Subordinationsrate zu Beginn der Studie Phasen, die sich durch eine fehlende Korrelation auszeichnen. In diesem Fall erstreckt sich diese Phase über die ersten Zeitpunkte und endet zwischen Zeitpunkt 5 und 6.

Auf den ersten Blick werden deutliche Unterschiede sichtbar. Während die gleitende Korrelation für TN024 hier u-förmig ist, findet sich für TN025 eine wellenförmige Entwicklung. Die Korrelation für die Daten von TN024 wird nach der angesprochenen Phase am Anfang erst positiv und dann negativ. In dieser Phase negativer Korrelation werden große Effektstärken sichtbar. Diese Entwicklung kehrt sich etwa ab Zeitpunkt 20 wieder um. Die Korrelationskoeffizienten tendieren zunächst gegen null und werden ab Zeitpunkt 22 wieder positiv. Im Gegensatz dazu zeigen die Korrelationskoeffizienten in den Daten von TN025 eine wellenförmige Bewegung, die sich sowohl durch große Effektstärken im positiven Bereich (etwa um Zeitpunkt 11) als auch im negativen Bereich (ab Zeitpunkt 21) auszeichnet.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).



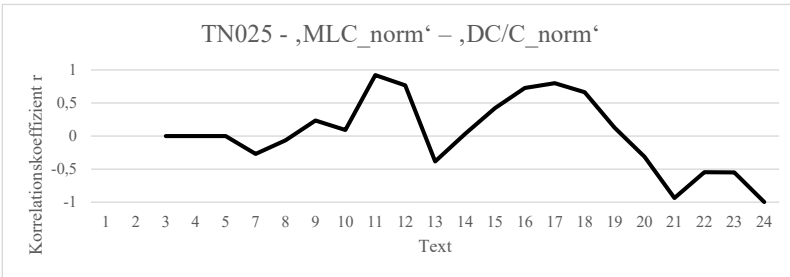


Abbildung 52: Gleitende Korrelation ‚MLC_norm' – ‚DC/C_norm' für TN024 (oben) und TN025 (unten)

In den generellen Korrelationsanalysen können hinsichtlich dieses Variablenpaares weder für TN024 noch für TN025 signifikante Korrelationen nachgewiesen werden (vgl. Tabelle 62 und Tabelle 63).

‚Durchschnittliche Länge einer MSE' (MLC) – ‚Rate komplexer NPs pro MSE'

In Abbildung 53 werden die gleitenden Korrelationen zwischen der Länge minimaler satzwertiger Einheiten und den komplexen Nominalphrasen pro MSE gezeigt. Im Gegensatz zu den oben vorgestellten Korrelationen (vgl. Abbildung 52) lassen sich hier prägnante Übereinstimmungen in den Daten von TN024 und TN025 herausarbeiten.

Für jede der gleitenden Korrelationen lässt sich konstatieren, dass diese zu Beginn der Zeitreihe negativ sind und zunächst gegen null tendieren. Zwischen Zeitpunkt 4 bzw. 6 werden die Korrelationskoeffizienten für beide TN positiv. Für TN024 bleiben sie das über große Teile des Erhebungszeitraumes und zeigen zum Teil sehr große Effektstärken. Dies wird zwischen Zeitpunkt 11 und Zeitpunkt 18 deutlich. Erst zum Ende der Studie verändert sich die Korrelation der beiden Variablen für TN024. Diese liegen zwischen Zeitpunkt 19 und 21 etwa bei null und werden erst zum Ende der Studie wieder positiv. Die Entwicklung der Korrelationskoeffizienten ist in den Daten von TN025 insgesamt moderater. Nach dem beschriebenen Anstieg zu Beginn der Studie nähern sich diese etwa

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

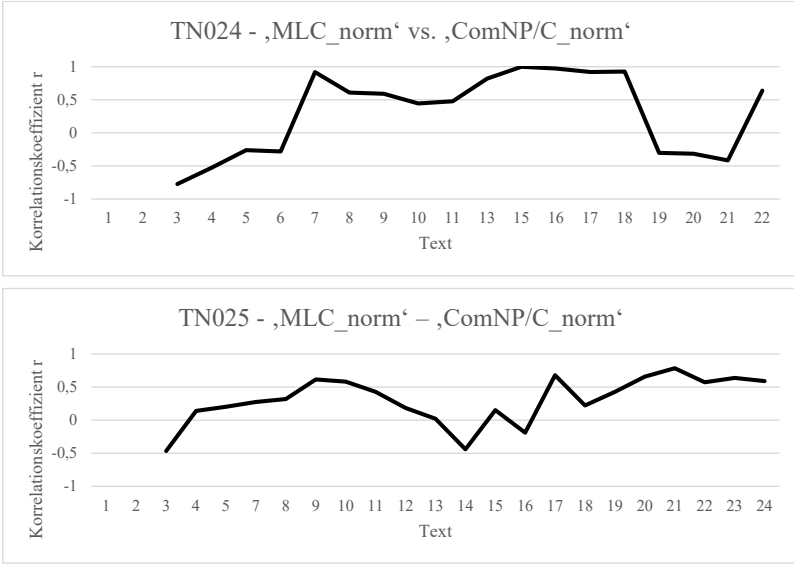


Abbildung 53: Gleitende Korrelation ‚MLC_norm' – ‚ComNP/C_norm' für TN024 (oben) und TN025 (unten)

ab Zeitpunkt 10 wieder null an. Erst ab Zeitpunkt 18 lassen sich größere Effektstärken beobachten, die jedoch nie die Größe der Effektstärken in den Daten von TN024 erreichen.

Für TN024 ergibt die generelle Korrelation dieser beiden Variablen eine signifikante Korrelation mit großer Effektstärke. Dieses Bild zeigen die gleitenden Korrelationen ebenfalls. Es wird deutlich, dass vor allem die Phase zwischen den Zeitpunkten 6 und 18 diese Korrelation beeinflusst. Für TN025 kann in diesem Fall keine signifikante Korrelation nachgewiesen werden.

‚Subordinationsrate' – ‚Rate komplexer NPs pro MSE'

Die letzte gleitende Korrelation in Bezug auf die ausgewählten Komplexitätsmaße ist die Korrelation zwischen der Subordinationsrate und den komplexen Nominalphrasen pro MSE (vgl. Abbildung 54). Neben der beschriebenen Anfangsphase, in der die Werte aufgrund der ‚Subordinationsrate' bei null liegen, unterscheiden sich die beiden gleitenden Korrelationen für TN024 und TN025 relativ deutlich. Die Entwicklung der Korrelationskoeffizienten der Daten von TN024 zeichnet sich durch relative Stabilität aus und liegt im Zeitverlauf der Studie oft in der Nähe von null. Dies spricht dafür, dass es für TN024 keine beobachtbare Korrelation zwischen der ‚Subordinationsrate' und der ‚Rate komplexer Nominalphrasen pro MSE' gibt. Für die meisten der beweglichen Fenster

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

ergeben sich in beiden gleitenden Korrelationen negative Werte. In der Datenreihe von TN025 zeigt sich eine Spitze zwischen Zeitpunkt 16 und 18, bei der die Korrelation sowohl positiv ist als auch eine große Effektstärke erreicht.

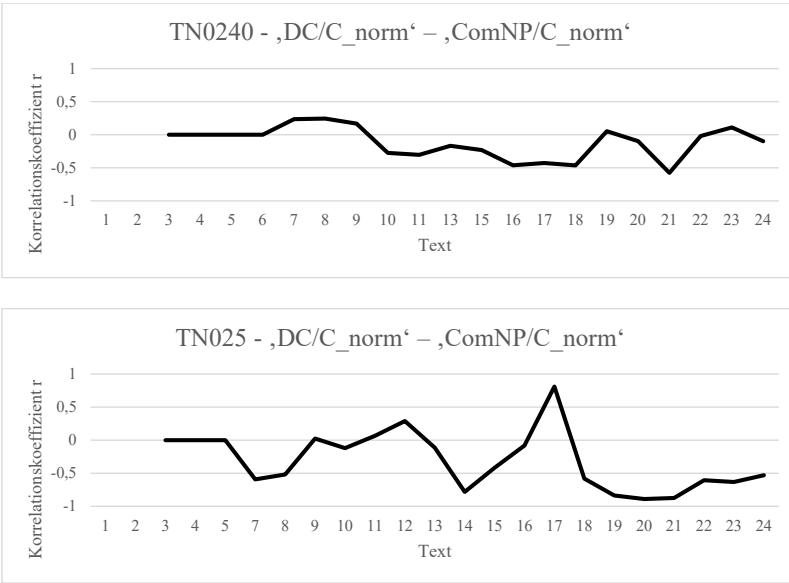


Abbildung 54: Gleitende Korrelation ‚DC/C_norm' – ‚ComNP/C_norm' für TN024 (oben) und TN025 (unten)

Es lässt sich an dieser Stelle zusammenfassend festhalten, dass vor allem die Korrelationen der verschiedenen Wortstellungsmuster die bisherigen Ergebnisse um Perspektiven ergänzen können, die zuvor in den Daten schwieriger zu identifizieren waren. Dies gilt ebenso für den Unterschied zwischen generellen und gleitenden Korrelationen. Bei der Berechnung der generellen Korrelationen der Syntaxvariablen bzw. der Komplexitätsmaße können hingegen nur wenige Zusammenhänge herausgearbeitet werden. Ein Ergebnis, dass sich bei beiden Teilnehmenden andeutet, war, dass längere satzwertige Einheiten positiv mit der Anzahl komplexer Nominalphrasen korrelieren. Zusammenhänge zwischen der Länge von satzwertigen Einheiten und der Rate von Verben in Endstellung oder ein Zusammenhang zwischen Verben in Endstellung und komplexen Nominalphrasen können nicht herausgearbeitet werden. Deutlich wird aber, dass in vielen Fällen Variablen deshalb nicht signifikant miteinander korrelieren, weil sich die Beziehung dieser Variablen im Zeitverlauf mehrfach ändert. Es kann gezeigt werden, dass gleitende Korrelationen eine Möglichkeit bieten, diese Veränderungen besser nachvollziehen zu können.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

In einem letzten Analyseschritt sollen nun die bisher gewonnenen Erkenntnisse vor dem Hintergrund möglicher signifikanter *Change Points* in den Blick genommen werden.

4.2.8 *Change Point*-Analysen

In den folgenden Abschnitten werden die Ergebnisse der *Change Point*-Analysen (CPAs) vorgestellt, die für die beiden Teilnehmenden TN024 und TN025 anhand der Variablen (‚Erreichte Profilstufe', ‚Durchschnittliche Länge MSE', ‚Subordinationsrate' und ‚Rate der komplexen Nominalphrasen pro MSE') durchgeführt werden (für eine Erklärung des methodischen Vorgehens bei einer CPA vgl. Kapitel 3.3.5).[219] Die schwarze Linie im grafischen Output einer CPA (vgl. z. B. Abbildung 55) verdeutlicht jeweils die Entwicklung der Variable über die Zeit. Die hellblauen Flächen im Hintergrund zeigen die Ergebnisse der CPA. Kommt es zu einer Verschiebung dieses hellblauen Bereichs, deutet dies auf die Existenz eines *Change Points* hin. Hier ist anzumerken, dass es in diesem Bereich bereits bei einer 50 %-Konfidenz für das Auftreten eines CPs zu einer Verschiebung des gezeigten Bereichs kommt. Als CP charakterisiert werden in der vorliegenden Studie nur solche Veränderungen, die mit mindestens 90 % Konfidenz identifiziert werden. Das heißt, dass in manchen der Schaubilder zwar Verschiebungen im hellblauen Bereich auftreten, aber trotzdem davon ausgegangen werden kann, dass es in den gezeigten Daten zu keiner systematischen Veränderung im Sinne der vorliegenden Studie kommt.

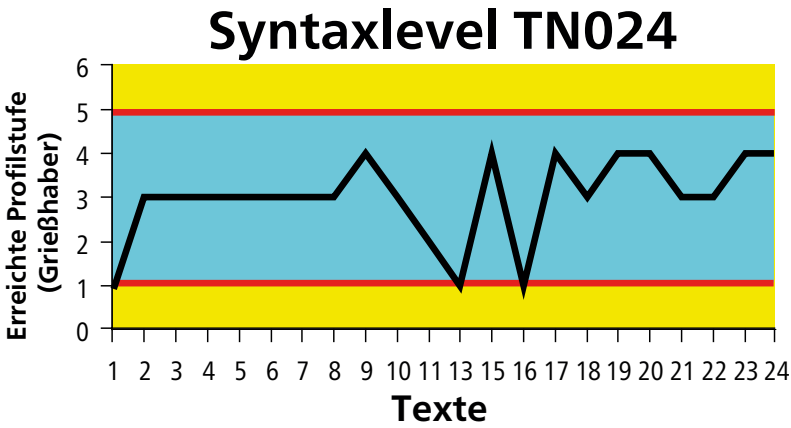
In manchen der im Folgenden gezeigten Plots sind darüber hinaus rote Linien, sogenannte Kontrolllimits der untersuchten Variablen, zu sehen. Diese Kontrolllimits repräsentieren die maximale Spannweite, innerhalb der die Werte der Variablen wahrscheinlich variieren werden. Da diese Kontrolllimits eine Normalverteilung der Daten annehmen, sind sie nicht für jedes der hier vorliegenden Datensets plausibel. Falls diese Kontrolllimits in den folgenden Schaubildern nicht zu sehen sind, liegt das an dem jeweils gewählten Ausschnitt auf der y-Achse, dessen Auswahl für jedes Schaubild kurz begründet wird. In den folgenden Ergebnisbeschreibungen werden die Output-Daten analysiert, die sich auf die Mittelwerte der beobachteten Variablen beziehen. Unter den Schaubildern befindet sich jeweils eine kurze Zusammenfassung der Analyseergebnisse als Teil des automatisierten Outputs einer CPA. Diese gibt, neben dem Ergebnis, ob CPs vorliegen, die berechneten Konfidenzintervalle, eine kurze Erklärung des verwendeten Bootstrapping-Verfahrens und eine Schätzung des Mittelwerts für die untersuchte Variable an.

219 Für eine ausführliche Beschreibung des Outputs einer CPA vgl. Kapitel 3.3.5 und ergänzend die Tutorials auf der Homepage des Entwicklers unter: https://variation.com/change-point-analyzer-tutorial-1/ [23.02.2023].

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

4.2.8.1 Change Point-Analyse: Erreichte Profilstufe

Als Erstes werden in den Abbildungen 56 und 57 die Ergebnisse der CPAs hinsichtlich der erreichten Profilstufen bzw. des Syntaxlevels für TN024 und TN025 vorgestellt.



Ergebnis: Keine signifikanten Veränderungen (CPs) für ‚Erreichte Profilstufe' gefunden

Confidence Level for Candidate Changes = 50%, Confidence Level for Inclusion in Table = 90%, Confidence Interval = 95%, Bootstraps = 1000, Without Replacement, MSE Estimates

Estimated Average = 3

Abbildung 55: Grafischer Output der CPA der erreichten Profilstufen (Profilanalyse nach Grießhaber) für TN024[220]

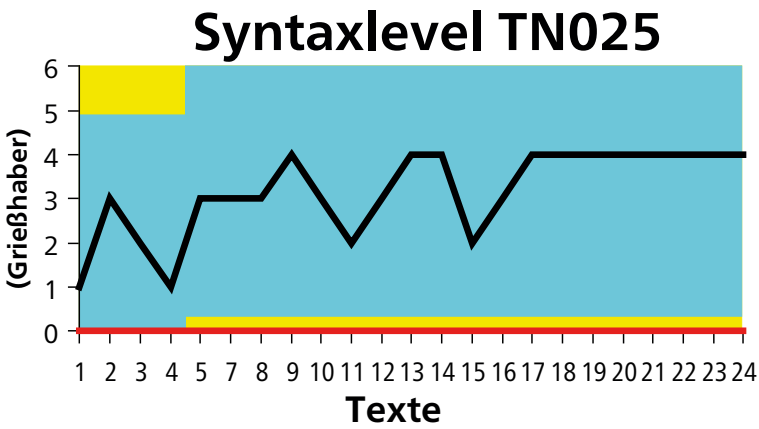
Für TN024 können in Bezug auf die jeweils erreichten Profilstufen über die gesamte Studiendauer keine CPs festgestellt werden (vgl. Abbildung 55). Ebenso können keine systematischen Veränderungen in der Varianz der Variablenwerte für TN024 ermittelt werden. In der Analyse werden keine Abweichungen von einer der zugrundeliegenden Annahmen festgestellt. D.h., es liegt eine unabhängige Fehlerstruktur innerhalb der Daten vor (vgl. Kapitel 3.3.5). Der geschätzte Durchschnitt der Werte für die Variable liegt bei 3. Die Variabilität zwischen den erreichten syntaktischen Profilstufen ist so hoch, dass keine systematischen Wechsel berechnet werden können. Es wird deutlich, dass sich Phasen relativer Stabilität mit Phasen deutlicher Variabilität abwechseln (vgl. zur Variabilität der

220 Auf der y-Achse sind die Profilstufen nach Grießhaber abgebildet und auf der x-Achse die einzelnen Texte (n=22), die TN024 geschrieben hat. Für TN024 liegen die Texte 12 und 14 nicht vor (vgl. das Kapitel 3.3.2 zur Auswahl der Texte).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Variablen Kapitel 4.2.2). Während die Phase zwischen Text 2 (5 Wochen / 98 UE) und Text 8 (17 Wochen / 307 UE) kaum Variabilität zeigt, ergeben sich zwischen Text 9 (18 Wochen / 347 UE) und Text 17 (42 Wochen / 697 UE) sehr unterschiedliche Werte für die Variable ‚Erreichte Profilstufe'.[221]

Für TN025 ergab die CPA in Bezug auf die erreichten Profilstufen folgendes Ergebnis:



Ergebnis: Tabelle signifikanter Veränderungen (CPs) für ‚Erreichte Profilstufe'

Confidence Level for Candidate Changes = 50%, Confidence Level for Inclusion in Table = 90%, Confidence Interval = 95%, Bootstraps = 1000, Without Replacement, MSE Estimates

Column A	Confidence Interval	Conf. Level	From	To	Level
5	(3, 8)	96%	1.75	3.4737	2

Abbildung 56: Grafischer Output der CPA der erreichten Profilstufen (Profilanalyse nach Grießhaber) für TN025[222]

221 Im Folgenden werden jeweils bei der ersten Nennung eines Textes bzw. eines Zeitpunktes die Wochen seit Beginn des Sprachkurses und die bis zu diesem Zeitpunkt besuchten Unterrichtsstunden angegeben.

222 Auf der y-Achse sind die Profilstufen nach Grießhaber abgebildet und auf der x-Achse die einzelnen Texte (n=23), die TN025 produziert hat. Für TN025 liegt Text 6 nicht vor (vgl. Kapitel 3.3.2 zur Auswahl der Texte).

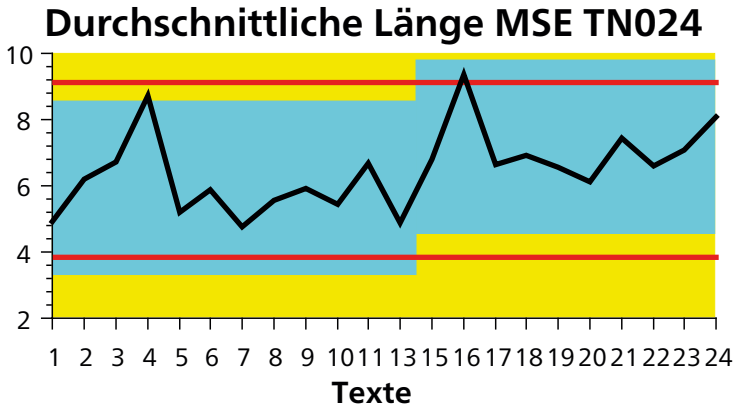
© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Für TN025 kann durch die CPA ein systematischer *Change Point* bei Text Nummer 5 ‚Regeln in Deutschland' (nach 10 Wochen / 205 UE) ermittelt werden. Dieser CP wird mit einer Konfidenz von 96 % ermittelt. Zusätzlich zeigt das Konfidenzintervall (KI 95 %), dass dieser CP zwischen Text 3 (7 Wochen / 110 UE) und Text 8 (17 Wochen / 300 UE) auftritt. Die durchschnittlich erreichte Profilstufe vor dem CP ist 1,75 und steigt nach diesem auf 3,4737. Außerdem wird der gefundene CP als Level 2-*Change Point* angegeben, das heißt, dass das Programm diesen erst bei einem zweiten Durchlauf durch die Daten entdeckte und er daher mit Vorsicht interpretiert werden sollte. Es werden keine Abweichungen von einer der zugrundeliegenden Annahmen festgestellt. Zeitpunkt 5 (10 Wochen / 205 UE) markiert in der Variabilitätsanalyse den Beginn einer zunehmenden Variabilität in den Werten der Variable ‚Erreichte Profilstufe', die dann etwa bis zum 17. Text (42 Wochen / 734 UE) anhält. Das errechnete Konfidenzintervall stimmt außerdem mit der ersten Phase der Variabilität in den Daten von TN025 überein, die zu Zeitpunkt 3 (7 Wochen / 110 UE) beginnt und zu den Zeitpunkten 5 (10 Wochen / 205 UE) und 7 (15 Wochen / 285 UE) die größte Spannweite erreicht. Text Nummer 8 (17 Wochen / 300 UE) markiert sowohl das Ende der ersten Phase der Variabilität als auch den Sprung in die zweite Phase.

4.2.8.2 *Change Point*-Analyse: Durchschnittliche Länge minimaler satzwertiger Einheiten

Die nächste untersuchte Variable ist für TN024 und TN025 die ‚Durchschnittliche Länge MSE'. Für TN024 ergibt sich folgendes Bild: Für die Variable ‚Durchschnittliche Länge MSE' ergibt die CPA eine systematische Veränderung der Zeitreihenwerte und zwar mit einer Konfidenz von 92 % zum Zeitpunkt, an dem Text 15 (35 Wochen / 573 UE) geschrieben wird (vgl. Abbildung 57). Das zeitliche Konfidenzintervall erstreckt sich relativ weit von Text 7 (15 Wochen / 292 UE) bis zu Text 23 (60 Wochen / 1010 UE). Eine MSE hat vor diesem Zeitpunkt eine durchschnittliche Länge von 5,9 und nach diesem Zeitpunkt von 7,1. Der gefundene CP wird als Level 1-*Change Point* charakterisiert. Es werden keine Abweichungen von einer der zugrundeliegenden Annahmen festgestellt. Wie in der Analyse zur erreichten Profilstufe bei TN025 fällt der gefundene CP in eine Phase sehr großer Variabilität innerhalb der Zeitreihenwerte der Variablen (vgl. Kapitel 4.2.2). Von Zeitpunkt 11 (25 Wochen / 481 UE) bis zu Zeitpunkt 18 ist die Variabilität in den Daten am größten. Wie zuvor fällt diese Phase vollständig in das durch die CPA berechnete Konfidenzintervall.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).



Ergebnis: Tabelle signifikanter Veränderungen (CPs) für ‚Durchschnittliche Länge MSE'

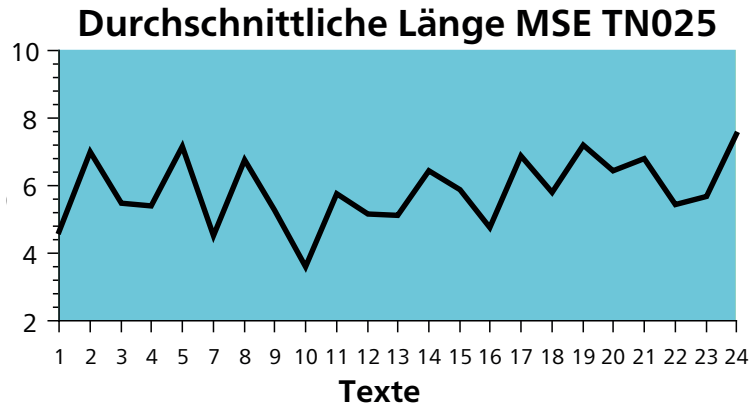
Confidence Level for Candidate Changes = 50%, Confidence Level for Inclusion in Table = 90%, Confidence Interval = 95%, Bootstraps = 1000, Without Replacement, MSE Estimates

Column A	Confidence Interval	Conf. Level	From	To	Level
15	(7, 23)	92%	5.886	7.134	1

Abbildung 57: Grafischer Output der CPA für die Variable ‚Durchschnittliche Länge MSE' (MLC) für TN024[223]

223 Der Ausschnitt auf der y-Achse der beiden Schaubilder zur durchschnittlichen Länge minimaler satzwertiger Einheiten bildet die Werte für die durchschnittliche Länge minimaler satzwertiger Einheiten in den geschriebenen Texten zwischen den Werten 2 und 10 ab. Der minimale Wert ergibt sich aus der minimalen Länge einer satzwertigen Einheit, die in der Regel ein Subjekt und ein Prädikat umfasst. Der maximale Wert ergibt sich aus den vorliegenden Daten der Geschwister, in denen dieser Wert in keinem der Texte überschritten wird. Zur x-Achse vgl. Fußnoten 220 und 222.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).



Ergebnis: Keine signifikanten Änderungen (CPs) für ‚Durchschnittliche Länge MSE' gefunden

Confidence Level for Candidate Changes = 50%, Confidence Level for Inclusion in Table = 90%, Confidence Interval = 95%, Bootstraps = 1000, Without Replacement, MSE Estimates

Estimated Average = 5.8398525

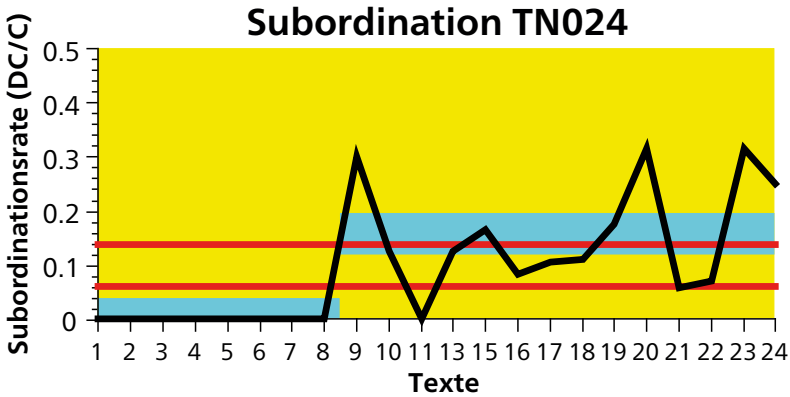
Abbildung 58: Grafischer Output der CPA für die Variable ‚Durchschnittliche Länge MSE' (MLC) für TN025

Für TN025 können im Gegensatz zu TN024 in Bezug auf die durchschnittliche Länge der MSE keine systematischen Veränderungen identifiziert werden (vgl. Abbildung 58). Die durchschnittliche Länge von MSE liegt über die gesamte Studiendauer bei etwa 5,8. Es werden keine Abweichungen von einer der zugrundeliegenden Annahmen festgestellt. Insgesamt scheint die Variable ‚Durchschnittliche Länge MSE' für TN025 sehr stabil zu sein. In den Daten zur Varianz der Variable werden keine signifikanten Änderungen gefunden. Diese Variable fällt darüber hinaus bei der Variabilitätsanalyse als stabile Variable mit geringer Variabilität auf, was der nicht erfolgten Identifizierung einer signifikanten Veränderung zusätzlich Plausibilität verleiht.

4.2.8.3 *Change Point*-Analyse: Subordination

Die nächsten beiden Schaubilder verdeutlichen die Ergebnisse der CPA hinsichtlich der ‚Subordinationsrate'.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).



Ergebnis: Tabelle signifikanter Veränderungen (CPs) für ‚Subordinationsrate'

Confidence Level for Candidate Changes = 50%, Confidence Level for Inclusion in Table = 90%, Confidence Interval = 95%, Bootstraps = 1000, Without Replacement, MSE Estimates

Column A	Confidence Interval	Conf. Level	From	To	Level
9	(9,17)	98%	0	0.15761	1

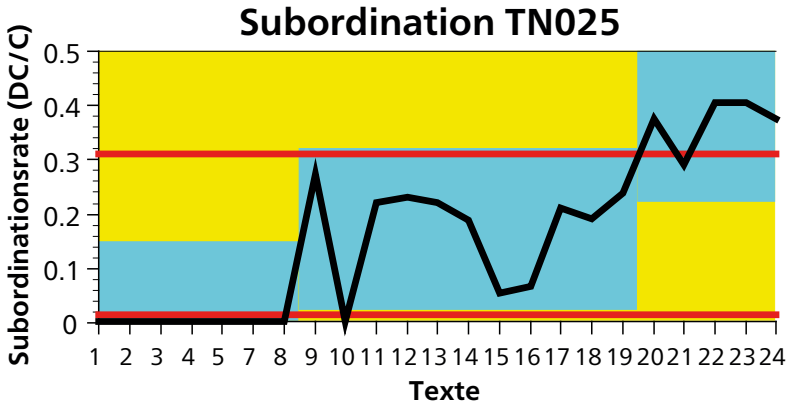
Abbildung 59: Grafischer Output der CPA für die Variable ‚Subordination' (DC/C) für TN024[224]

In Bezug auf das Verhältnis von Hauptsätzen und unterordnenden (oder subordinierenden) Nebensätzen kann bei TN024 eine systematische Veränderung im Zeitverlauf festgestellt werden (vgl. Abbildung 59). Diese tritt mit einer Konfidenz von 98 % zum Zeitpunkt von Text 9 (18 Wochen / 347 UE) auf. Das zeitliche Konfidenzintervall (KI 95 %) umfasst die Zeitpunkte zwischen Text 9 und Text 17 (42 Wochen / 697 UE). Die Subordinationsrate beträgt vor Text 9 im Schnitt 0 und danach 0,15761. Der gefundene CP wird als Level 1-*Change Point* charakterisiert und es werden keine Abweichungen von einer der zugrundeliegenden Annahmen festgestellt. Die Analyse der Varianz der Daten ergibt keine signifikanten Veränderungen in den Daten.

224 Der Ausschnitt auf der y-Achse der beiden Schaubilder zur ‚Subordinationsrate' ist von 0 bis 0,5 skaliert. Der minimale Wert ergibt sich aus dem potenziell niedrigsten Wert dieser Rate, falls keine Nebensätze mit Verbendstellung produziert werden. Der maximale Wert ergibt sich aus den vorliegenden Daten der beiden Lernenden, in denen dieser Wert in keinem der Texte überschritten wird. Zur x-Achse vgl. Fußnoten 220 und 222.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Analog zu den bisher gefundenen CPs liegt Zeitpunkt 9 in der Mitte der ersten Phase mit größerer Variabilität. In den min-max-Grafiken ist zu Zeitpunkt 9 die größte während der gesamten Studiendauer beobachtete Variabilität festzustellen.



Ergebnis: Tabelle signifikanter Veränderungen (CPs) für ‚Subordinationsrate'

Confidence Level for Candidate Changes = 50%, Confidence Level for Inclusion in Table = 90%, Confidence Interval = 95%, Bootstraps = 1000, Without Replacement, MSE Estimates

Column A	Confidence Interval	Conf. Level	From	To	Level
9	(9, 12)	98%	0	0.17176	2
20	(19, 20)	100%	0.17176	0.37066	1

Abbildung 60: Grafischer Output der CPA für die Variable ‚Subordination' (DC/C) für TN025

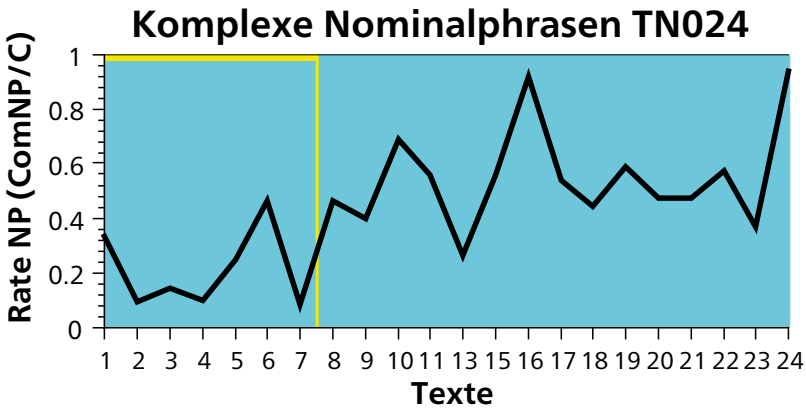
Für TN025 werden in den Werten der Variable ‚Subordinationsrate' insgesamt zwei systematische Veränderungen in der Datenreihe im Verlauf der Studie festgestellt (vgl. Abbildung 60). Der erste CP tritt mit einer 98 % Konfidenz zu Text 9 (18 Wochen / 338 UE) auf, wobei das zeitliche Konfidenzintervall (KI 95 %) sich von Text 9 zu Text 12 (28 Wochen / 558 UE) erstreckt. Die durchschnittliche Subordinationsrate beträgt vor dem Zeitpunkt 0 und danach 0,17176. Der gefundene CP wird als Level 1-*Change Point* charakterisiert. In diesem Fall wird Zeitpunkt 9 für beide Teilnehmende als ein Punkt signifikanter Veränderung identifiziert. Das Konfidenzintervall ist für TN025 etwas kleiner als für TN024. In der Variabilitätsanalyse liegt Zeitpunkt 9 in der Phase der größten Variabilität.

Außerdem wird mit einer Konfidenz von 100 % ein zweiter CP zum Zeitpunkt von Text 20 (45 Wochen / 808 UE) identifiziert. Dieser CP weist außerdem ein vergleichsweise enges zeitliches Konfidenzintervall (KI 95 %) auf, welches sich

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

nur zwischen den Texten 19 (45 Wochen / 803 UE) und 20 erstreckt. Dieser zweite CP wird als Level 1-*Change Point* identifiziert. Hinsichtlich der Variabilitätsanalyse liegt dieser Change Point innerhalb eines ‚steigenden Kanals', in dem die Variabilität konstant groß bleibt, während die Minimum- und die Maximum-Grenzen konstant steigen. Es werden keine Abweichungen von einer der zugrundeliegenden Annahmen festgestellt und die Analyse der Varianz der Daten ergab keine signifikanten Veränderungen in den Zeitreihendaten.

4.2.8.4 *Change Point*-Analyse: Komplexe Nominalphrasen



Ergebnis: Tabelle signifikanter Veränderungen (CPs) für ‚Komplexe Nominalphrasen pro MSE'

Confidence Level for Candidate Changes = 50%, Confidence Level for Inclusion in Table = 90%, Confidence Interval = 95%, Bootstraps = 1000, Without Replacement, MSE Estimates

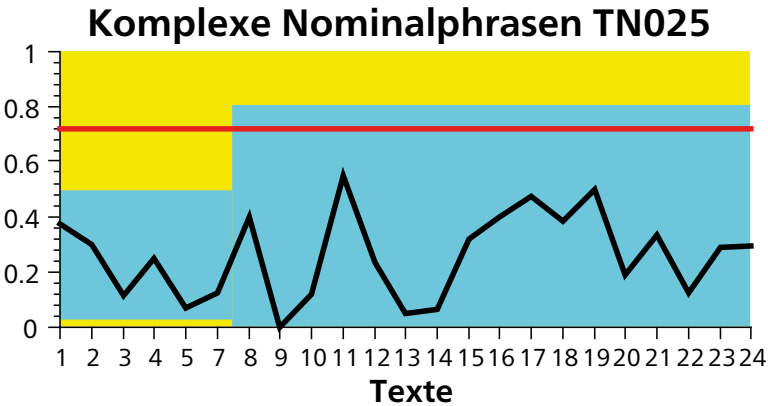
Column A	Confidence Interval	Conf. Level	From	To	Level
8	(6, 11)	98%	0.20885	0.54867	1

Abbildung 61: Grafischer Output der CPA für die Variable ‚Komplexe Nominalphrasen' (ComNP/C) für TN024[225]

225 Der Ausschnitt auf der y-Achse der beiden Schaubilder zur ‚Rate komplexer Nominalphrasen pro MSE' ist von 0 bis 1 skaliert. Der minimale Wert ergibt sich aus dem potenziell niedrigsten Wert dieser Rate, falls keine komplexen Nominalphrasen produziert werden. Der maximale Wert ergibt sich aus den vorliegenden Daten der beiden Lernenden, in denen dieser Wert in keinem der Texte überschritten wird. Zur x-Achse vgl. Fußnoten 220 und 222.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Bei der Analyse der Daten von TN024 in Bezug auf die Variable ‚Komplexe Nominalphrasen pro MSE' kann eine systematische Veränderung in der Zeitreihe festgestellt werden (vgl. Abbildung 61). Diese tritt mit einer Konfidenz von 98 % zum Zeitpunkt von Text 8 (17 Wochen / 307 UE) auf und das zeitliche Konfidenzintervall (KI 95 %) erstreckt sich von Text 6 (12 Wochen / 231 UE) bis zu Text 11 (25 Wochen / 481 UE). Die durchschnittliche Rate von komplexen Nominalphrasen pro MSE beträgt vor diesem Zeitpunkt 0,20885 und danach 0,54867. Der gefundene CP wird als Level 1-*Change Point* charakterisiert. Außerdem werden keine Abweichungen von einer der zugrundeliegenden Annahmen festgestellt. Die Analyse der Varianz der Daten ergibt keine zusätzlichen systematischen Veränderungen. Wie bei den vorangegangenen Analysen markiert der hier gefundene CP den Punkt der größten Variabilität innerhalb der Daten (vgl. dazu Kapitel 4.2.2). Es kommt in den Daten zu weiteren Phasen mit ähnlich großer Variabilität, in denen kein CP gefunden wird.



Ergebnis: Keine signifikanten Veränderungen (CPs) für ‚Komplexe Nominalphrasen pro MSE'

Confidence Level for Candidate Changes = 50%, Confidence Level for Inclusion in Table = 90%, Confidence Interval = 95%, Bootstraps = 1000, Without Replacement, MSE Estimates

Estimated Average = 0.25778989

Abbildung 62: Grafischer Output der CPA für die Variable ‚Komplexe Nominalphrasen' (ComNP/C) für TN025

Im Schaubild der Daten der Variable ‚Komplexe Nominalphrasen pro MSE' des TN025 deutet die Verschiebung der hellblau unterlegten Fläche zum Zeitpunkt des Textes 8 auf eine systematische Veränderung der Daten in der Zeitreihe hin (vgl. Abbildung 62). Dieser potenzielle CP erfüllt nicht die Voraussetzung einer 90 %igen Konfidenz. Für die vorliegenden Daten von TN025 kann daher nicht mit Sicherheit eine systematische Veränderung angenommen werden. Der

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

geschätzte Mittelwert der Variable beträgt 0,258. Es werden keine Abweichungen von einer der zugrundeliegenden Annahmen festgestellt.

Nicht in allen der untersuchten Zeitreihen können *Change Points* identifiziert werden. Diese könnte mit der zu großen Variabilität innerhalb der Zeitreihendaten zusammenhängen. In den Fällen, in denen CPs gefunden werden, fallen diese ausnahmslos in Phasen relativ großer Variabilität. Diese kann zuvor in den Variabilitätsanalysen herausgearbeitet werden. Während für alle CPs zutrifft, dass sie mit Phasen großer Variabilität verknüpft werden können, werden nicht in allen solchen Phasen CPs gefunden.

4.2.9 Zusammenfassung: Einzelfallanalysen

Ziel der Einzelfallanalysen war es individuelle Erwerbs- und Entwicklungsverläufe so genau wie möglich zu rekonstruieren und dabei so detailliert wie möglich zu modellieren. Dabei wurden die Daten von zwei sehr ähnlichen Lernenden anhand von Zeitreihendaten und unter Anwendung verschiedener Verfahren analysiert. Die Ähnlichkeit der Lernenden sollte eine größtmögliche Vergleichbarkeit ermöglichen und der multimethodische Ansatz dabei helfen eine Vielzahl von Facetten der zweitsprachlichen Entwicklung greifbar zu machen. Die Erkenntnisse darüber, wie ausgewählte syntaktische Strukturen sich in schriftlichen Lernertexten entwickeln und dabei miteinander zusammenhängen, kann letztlich dazu beitragen, Modelle der zweitsprachlichen Schreibentwicklung zu entwickeln bzw. weiterzuentwickeln.

Wie bei den Daten auf Gruppenebene deuteten die deskriptiven Daten für die beiden Lernenden zunächst auf Zusammenhänge zwischen zeitlichem Ablauf und Zunahme der untersuchten Variablen hin. In der zweiten Studienphase beinhalteten die Texte mehr Wörter und mehr minimale satzwertige Einheiten als in der ersten Studienphase. Die Werte der Raten, die die Subordination und die Verwendung komplexer Nominalphrasen verdeutlichen, erhöhten sich stark. Der geringste Unterschied zwischen beiden Studienphasen konnte im Komplexitätsmaß ‚Durchschnittliche Länge einer minimalen satzwertigen Einheit' beobachtet werden. Gleichzeitig wurden im Zeitverlauf konstant immer höhere Profilstufen erreicht, bis diese für beide Lernende ein Deckenniveau auf Profilstufe 4 ‚Verbendstellung' erreichten. Es ergaben sich Unterschiede zwischen TN024 und TN205. TN025 schrieb in beiden Studienphasen längere Texte und produzierte mehr MSE als TN024. Dabei schrieb TN024 längere minimale satzwertige Einheiten als TN025. Während TN024 längere satzwertige Einheiten produzierte, wurden diese von TN025 frequenter mit satzwertigen Einheiten, die ein Verb in Endstellung beinhalten, verbunden. Die Lernenden unterschieden sich bei der Integration komplexer Nominalphrasen in die produzierten Texte. Es fanden sich in den Texten von TN024 in beiden Studienphasen deutlich mehr komplexe NPs pro MSE als in denen von TN025. Die herausgearbeiteten Unterschiede könnten

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Hinweise darauf geben, wie unterschiedlich Lernende, die sich oberflächlich sehr ähnlich sind, Aufgaben bearbeiten und textuelle Strukturen realisieren.

Komplizierter war die Frage zu beantworten, ob sich aus den Daten der beiden Lernenden TN024 und TN025 Entwicklungssequenzen oder beschreibbare Stufen ableiten lassen. Mithilfe der in dieser Studie gewählten analytischen Verfahren waren Aussagen zur Implikationalität der beobachteten Lernverläufe nur indirekt möglich. Es wurde beispielsweise keine Implikationsanalyse unter Zuhilfenahme einer Implikationstabelle durchgeführt. Dennoch lassen sich hier wesentliche Erkenntnisse aus den in den Interaktions- und Trendanalysen beobachteten Daten anführen. Eindeutig war, dass für beide Lernende Stufe 4 ‚Verbendstellung' nach den Stufen 1–3 emergent wurde. Darüber hinaus ließen sich nur sehr vereinzelt Beispiele für die Stufen 5 und 6 nach Grießhaber in den Daten finden (vgl. zu diesem Punkt Kapitel 5). Die Emergenz der Stufen 1, 2 und 3 war kaum voneinander zu differenzieren. Vielmehr wurden alle drei Stufen zu sehr frühen Zeitpunkten, nämlich in den ersten Texten innerhalb der ersten Wochen, emergent. Zum Teil wurden mehrere Stufen zum selben Zeitpunkt emergent. Eine simultane Emergenz verschiedener Stufen wird von der *Processability Theory* jedoch ausgeschlossen (vgl. für eine ausführlichere Diskussion dieses Punktes Kapitel 5). Bei der Anwendung der Profilanalyse und Grießhabers Emergenzkriterium von drei Vorkommen pro Text zeigte sich außerdem, dass für TN024 Inversionen und Nebensätze zu einem früheren Zeitpunkt als Satzteile mit Separationen als erworben nachgewiesen werden konnten. Diese Punkte konstituieren prinzipiell keine negative Evidenz für die Existenz zugrundeliegender Erwerbssequenzen. Es könnte beispielsweise so sein, dass Strukturen zwar prinzipiell zur Verfügung standen bzw. als erworben gelten könnten, aber von den Lernenden nicht produziert wurden. Diese Punkte verdeutlichen die sehr hohen Anforderungen, die lernersprachliche Daten theoretisch erfüllen müssten, um tatsächlich Aussagen über implikationelle bzw. sequenzielle Entwicklungen treffen zu können (vgl. Kapitel 5).

Die Interaktions- und Trendanalysen verdeutlichten ebenfalls die große Variabilität innerhalb und zwischen den Daten der beiden Lernenden, die die beobachtete Entwicklung beider wesentlich auszeichnete. Die Trendanalysen ergänzten die deskriptiven, statistischen Verfahren in diesem Punkt und erlaubten vor allem in Verbindung mit den *Change Point*-Analysen die Interpretation von Facetten der Daten, die sonst nur schwer zugänglich wären (siehe unten). So konnten Unterschiede und musterhafte Gemeinsamkeiten in den Daten der beiden Geschwister aufgedeckt werden (vgl. für eine Diskussion, inwiefern Unterschiede mit persönlichen Schreibpräferenzen zusammenhängen könnten Kapitel 5). Die beobachteten Muster deuteten beispielsweise daraufhin, dass die Entwicklung von TN024 insgesamt variabler verlief als die von TN025, die sich hingegen durch längere stabile Phasen und eine insgesamt gleichmäßigere Entwicklung auszeichnete. Um diese Hypothesen zu überprüfen, müssten die bisherigen Analysen in einem nächsten Schritt beispielsweise durch Monte

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Carlo-Analysen ergänzt werden, die prüfen, ob die Zeitreihendaten von TN024 tatsächlich variabler sind als die von TN025.

Die durchgeführten Korrelationsanalysen verdeutlichten bekannte Zusammenhänge, wie die Korrelation von längeren, satzwertigen Einheiten und der steigenden Anzahl komplexer Nominalphrasen pro MSE. Insgesamt zeigten nur wenige der generellen Korrelationsanalysen statistisch signifikante Korrelationen. Hier konnten die gleitenden Korrelationen aufzeigen, warum dies für viele Korrelationen der Fall war, da sie visualisierten, inwiefern sich die Beziehungen verschiedener Variablen im Zeitverlauf immer wieder änderten und nicht konstant blieben. So vertieften gleitende Korrelationen das Verständnis der (unterschiedlichen) zweitsprachlichen Entwicklung von TN024 und TN025 weiter.

Zuletzt wurden mithilfe der *Change Point*-Analysen signifikante Veränderungen in den Zeitreihendaten der Lernenden identifiziert. Dies gelang nicht in allen Zeitreihen, was darauf schließen lässt, dass sich Zeitreihen ohne identifizierbare *Change Points* durch eine zu große Variabilität auszeichnen. Für TN024 konnten in der Entwicklung der drei Komplexitätsmaße, nicht aber in der Zeitreihe zu den erreichten Profilstufen, *Change Points* nachgewiesen werden. Für TN025 fand sich eine systematische Veränderung in der Zeitreihe zu den Profilstufen und außerdem *Change Points* in den Zeitreihenwerten der Variablen ‚Subordinationsrate' und ‚Rate komplexer Nominalphrasen pro MSE'. Keine systematische Veränderung konnte sich für TN025 in den Daten zur durchschnittlichen Länge einer MSE ermitteln lassen. Es konnte nur ein für beide Lernende übereinstimmender *Change Point* ermittelt werden. Dieser wurde in den Zeitreihen der Subordinationsrate zu Zeitpunkt 9 gefunden, was durchaus Sinn ergibt, da dies der erste Zeitpunkt war, zu dem Nebensatzstrukturen mit Verbendstellung erfolgreich produziert wurden.

Zusammenfassend lässt sich sagen, dass systematische Veränderungen einer Zeitreihe in den untersuchten Daten dann gefunden wurden, wenn der Zeitpunkt in eine Phase relativ großer Variabilität fiel. Immer, wenn ein *Change Point* gefunden wurde, ging dies mit sehr großer Variabilität einher, aber nicht jede Phase großer Variabilität hing mit einem *Change Point* zusammen. In diesen Fällen konnten über die verschiedenen Analyseschritte Variabilität und *Change Point*s miteinander in Beziehung gesetzt werden. Es schienen Zusammenhänge zwischen beiden Phänomenen zu existieren, die sehr unterschiedlicher Natur sein könnten (vgl. Kapitel 5).

Im nächsten Kapitel wurden die wesentlichen Ergebnisse der vorliegenden Arbeit im Hinblick auf die Beantwortung der Forschungsfragen reflektiert.

4.3 Zusammenfassung der Ergebnisse

In dieser Arbeit wurde das Ziel verfolgt, auf der Grundlage von schriftlichen Lernertexten die drei folgenden Forschungsfragen zu beantworten:

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Erwerbsreihenfolgen: Welche syntaktische(n) Erwerbsreihenfolge(n) ergibt (ergeben) sich für erwachsene Deutschlernende mit Arabisch als L1? (Forschungsfrage 1)

Entwicklung: Wie lässt sich die syntaktische Entwicklung der gesamten Gruppe und von ausgewählten Teilnehmenden charakterisieren? (Forschungsfrage 2)

Instruktion: Lassen sich aufgrund der Instruktion bzw. der experimentellen Manipulation innerhalb der Studie (explizite Sprachvermittlung mit Schwerpunkt auf Syntax vs. implizite Vermittlung mit Schwerpunkt auf Semantik) Unterschiede zwischen den Teilnehmerinnen und Teilnehmern feststellen? (Forschungsfrage 3)

Die ersten beiden Fragen bezogen sich darauf, anhand der vorliegenden lernersprachlichen Daten zweitsprachliche Entwicklungsprozesse sichtbar zu machen und zu überprüfen, inwiefern die Lernerinnen und Lerner in ihrer Entwicklung bestimmten Erwerbsreihenfolgen bzw. Erwerbssequenzen folgen. Dabei wurden zwei verschiedene longitudinale Untersuchungsdesigns ausgewählt. Zum einen wurden Daten von 44 Lernenden zu drei Zeitpunkten innerhalb der ersten sechs Monate des Zweitspracherwerbs ausgewählt. Zum anderen wurde die zweitsprachliche Entwicklung von zwei Lernenden anhand der Daten zu 24 Zeitpunkten und über insgesamt 18 Monate analysiert. Die dritte Frage bezog sich auf mögliche Effekte, die sich aus der experimentellen Manipulation des Unterrichts ergeben könnten. Hier wurden mehrere ANOVAs mit Messwiederholung durchgeführt, die in den Daten der 44 Lernenden die Lernerinnen und Lerner aus den beiden Instruktionsgruppen miteinander in Beziehung setzten. Zur Überprüfung möglicher Zusammenhänge zwischen Experimentalgruppe und Zugehörigkeit zu einem der longitudinalen Cluster wurde ein Chi Quadrat-Test durchgeführt.

Zur Beantwortung von Forschungsfrage 1 wurden in einem ersten Schritt die deskriptiven Daten der 44 Teilnehmenden hinsichtlich der zu den drei Untersuchungszeitpunkten produzierten minimalen satzwertigen Einheiten und den in den Texten erreichten Profilstufen analysiert. Diese verdeutlichten, dass von Zeitpunkt zu Zeitpunkt längere Texte geschrieben wurden, die mehr und längere minimale satzwertige Einheiten beinhalteten. Die deskriptiven Daten zu den Profilanalysen zeigten, dass relativ früh in der (gesteuerten) zweitsprachlichen Entwicklung ‚komplexe' Strukturen der Profilstufen 3 und 4 produziert wurden. Zu Zeitpunkt 2 machten diese Strukturen etwa 30 % aller minimalen satzwertigen Einheiten aus. Dies führte dazu, dass die Texte als Ganzes diesen Profilstufen zugeordnet werden konnten und zu Zeitpunkt 2 die Hälfte der Texte den komplexeren Stufen 3 und 4 zugeordnet wurden. Zu Zeitpunkt 3 wurden mehr als 90 % der Texte auf diesen beiden Profilstufen eingeordnet. Dass die Streuung innerhalb der Stichprobe mit jedem Zeitpunkt abnahm und zu Zeitpunkt 3 am geringsten war, verdeutlicht, dass die Teilnehmenden hinsichtlich der erreichten Profilstufen konvergieren und sich mit Profilstufe 4 zunächst ein

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Deckenniveau einstellt. Als nächstes ergab ein Friedman-Test, der überprüft, ob sich die ermittelten Profilstufen für die 44 Lernenden über die drei Zeitpunkte signifikant ändern, dass es zwischen allen Zeitpunkten jeweils zu statistisch signifikanten Änderungen kommt. Dies wurde in Kombination mit den aus den deskriptiven Statistiken abgeleiteten Tendenzen als Beleg dafür gewertet, dass sich aus den Gruppendaten ein möglicherweise sequenzieller Erwerbsverlauf ableiten lässt, in dem im Zeitverlauf immer höhere Erwerbs- bzw. Profilstufen erreicht werden. Um diese Gruppendaten mit den Daten einzelner Lernender zu vergleichen, wurden daran anschließend mit den Daten der 44 Lernenden Emergenzanalysen durchgeführt. Hierbei wurden zwei Emergenzkriterien nach Pienemann (1998) und Grießhaber (2012) vergleichend gegenübergestellt. Diese Analysen verdeutlichen, dass verschiedene Emergenzkriterien in Bezug auf dieselben lernersprachlichen Daten unterschiedliche Erwerbsverläufe evozieren. Ähnliche Tendenzen arbeitet bei der Kombination unterschiedlicher Kriterien ebenfalls Wisniewski (2020) heraus. Außerdem zeigte sich, dass gerade bei den hier vorliegenden Texten zu Beginn der zweitsprachlichen Schreibentwicklung bei der Anwendung des Emergenzkriterium nach Grießhaber deutlich mehr Texte in die Analyse eingeschlossen werden konnten, die bei der Anwendung des Emergenzkriteriums nach Pienemann aufgrund nicht ausreichender Belegstellen aus den Analysen ausgeschlossen werden mussten. Ersichtlich wurde, dass zunächst uneindeutige und auf einen Zeitpunkt bezogene Erwerbssequenzen eindeutiger werden, wenn die Analyse auf alle drei Zeitpunkte gleichzeitig bezogen wurde (vgl. zu den Implikationen dieser Ergebnisse die Diskussion in Kapitel 5).

Daran anschließend wurden in einer Einzelfallstudie die Texte von zwei Lernenden zu 24 Zeitpunkten analysiert. Hier bildeten die minimalen satzwertigen Einheiten und die ihnen jeweils zugeordneten Profilstufen die Grundlage der Analyse. Mit diesen Daten und den in dieser Studie ausgewählten Methoden ließen sich für die beiden Lernenden TN024 und TN025 keine eindeutigen Erwerbssequenzen rekonstruieren. Dies bezieht sich sowohl auf die Implikationalität als auch auf den sequenziellen Charakter der Profilstufen und kann damit begründet werden, dass in den Texten der beiden Lernenden verschiedene Profilstufen erstmals zu denselben Zeitpunkten emergent wurden. TN024 und TN025 produzierten in den ersten Texten innerhalb des Erhebungszeitraumes und zu Beginn ihrer zweitsprachlichen Entwicklung Strukturen der Profilstufen 1, 2 und 3. Nur für die Profilstufe 4 ließ sich eindeutig eine zeitlich nachgeordnete Emergenz feststellen. Es ließ sich allerdings auch keine negative Evidenz in Bezug auf die der Entwicklung zugrundeliegenden Erwerbssequenzen herausarbeiten.

Zur Beantwortung von Forschungsfrage 2 wurde auf Gruppenebene mithilfe longitudinaler Clusteranalysen untersucht, ob sich anhand der ausgewählten Komplexitätsmaße Gruppen von Lernenden identifizieren lassen, die sich in ihrer Entwicklung über die drei Zeitpunkte ähneln. Zusätzlich wurden in Einzelfallanalysen (grafische) Trendanalysen, unterschiedliche Korrelationsanaly-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

sen und *Change Point*-Analysen eingesetzt. Im Gegensatz zu Forschungsfrage 1 wurden zur Beantwortung von Forschungsfrage 2 zusätzlich zu den Parametern aus den Profilanalysen syntaktische Komplexitätsmaße herangezogen. Die syntaktische Entwicklung wurde anhand von drei unterschiedlich granularen Parametern verfolgt: ‚Durchschnittliche Länge einer minimalen satzwertigen Einheit', ‚Subordinationsrate pro MSE' und ‚Rate komplexer Nominalphrasen pro MSE'. Über die Clusteranalysen konnten anhand dieser Komplexitätsmaße drei Cluster identifiziert werden, die sich zu den Zeitpunkten 1 und 2 vor allem hinsichtlich der Länge der satzwertigen Einheiten unterscheiden. Zu Zeitpunkt 3 wurden zusätzlich die Werte der Rate komplexer Nominalphrasen pro MSE zu einem distinktiven Clustermerkmal. Die dritte Clustervariable ‚Subordinationsrate' schien zu keinem der drei Zeitpunkte ein wesentlicher Faktor bei der Zuordnung zu einem Cluster zu sein. Im Folgenden wurden Teilnehmende anhand der Anzahl ihrer Clusterwechsel über die drei Zeitpunkte in sieben verschiedene longitudinale Entwicklungscluster gruppiert, die sich vor allem aufgrund der Variabilität der ihnen zugrundeliegenden Daten unterscheiden. Dies wurde mithilfe von Datenvisualisierungen weiter überprüft.

In den Einzelfallanalysen anhand der Daten von TN024 und TN25 konnten durch die grafischen Trendanalysen zunächst variablere und stabilere Phasen der syntaktischen Entwicklung in den Zeitreihendaten der beiden identifiziert werden. Zusammengenommen mit den *Change Point*-Analysen zeigte sich, dass alle systematischen Veränderungen in den Zeitreihendaten in Phasen großer Variabilität fallen. Es kam jedoch nicht in jeder durch größere Variabilität geprägten Phase zu einer solchen Veränderung. In Phasen, in denen die untersuchten Variablen relativ stabil sind, konnten ebenfalls keine *Change Points* identifiziert werden. Es konnten zudem nicht in allen Zeitreihen der beiden Lernenden signifikante Veränderungen festgestellt werden. In der Zeitreihe zu den erreichten Profilstufen ergab sich für TN024 kein und für TN025 ein *Change Point*. Falls keine signifikanten Veränderungen innerhalb einer Zeitreihe identifizierbar waren, wird dies in der vorliegenden Studie als Hinweis darauf gesehen, dass die Zeitreihendaten als Ganzes zu stark durch Variabilität geprägt sind (vgl. zu den Implikationen dieses Punktes Kapitel 5). Die Interaktionsanalysen und die Analyse mit gleitenden Korrelationen verdeutlichten, dass sich die Beziehungen zwischen den beobachteten Variablen innerhalb der Zeitreihendaten für alle Variablen mehrfach ändern und nur selten in stabilere Phasen eintreten (vgl. zu diesem Punkt die Diskussion im nächsten Punkt). Dabei ergaben sich für TN024 und TN205 jeweils sehr individuelle Muster. Durch diese Verfahren konnten darüber hinaus auffällige Texte in der Zeitreihe entdeckt und genauer analysiert werden, die über die Auswertung der Rohdaten und der deskriptiven Statistik innerhalb der Zeitreihe bis dahin nicht weiter in den Fokus gerückt waren. Auffällig war, dass die Daten von TN025 sich insgesamt durch scheinbar stabilere Entwicklungen auszeichneten, während in den Daten von TN024 eine größere Variabilität zu beobachten war, die sich zum Zeitpunkt des Goethe-Zertifikats B1 allerdings abrupt verringerte.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

In Bezug auf Forschungsfrage 2 lässt sich zusammenfassend festhalten, dass die Kombination der ausgewählten methodischen Verfahren auf verschiedenen Ebenen eine Reihe von unterschiedlichen Facetten der zweitsprachlichen Entwicklung sichtbar machen konnte. Dabei wurde deutlich, dass die zweitsprachliche syntaktische Entwicklung nicht linear verläuft und Variabilität als relevantes Entwicklungsphänomen Aufschluss über eine Reihe von Prozessen geben kann bzw. unter Umständen mit für diese verantwortlich ist (vgl. zu diesem Punkt außerdem Kapitel 5). Gerade die Einzelfallanalysen zeigten, dass die beobachtbare lernersprachliche Entwicklung syntaktischer Strukturen stark von den Kontexten, in denen die Texte produziert wurden, und von den in diesen Kontexten elizitieren Textsorten abhängt. Gleichzeitig liefen diese Entwicklungen der lernersprachlichen Syntax innerhalb von bestimmten unter Umständen systeminternen Grenzen ab, die in der Diskussion im nächsten Kapitel genauer besprochen werden.

Zur Beantwortung von Forschungsfrage 3 zur experimentellen Manipulation bzw. zur Unterrichtsintervention innerhalb der Studie wurden mit den drei verwendeten syntaktischen Komplexitätsmaßen zunächst zweifache ANOVAs mit Messwiederholung durchgeführt. Diese wurden durch einen Chi Quadrat-Test in Bezug auf mögliche Zusammenhänge zwischen den longitudinalen Clustern und der Experimentalgruppe ergänzt. Es lässt sich konstatieren, dass anhand der hier untersuchten Variablen kein Effekt dieser Manipulation erkennbar war. Weder bei den ANOVAs noch bei dem Chi Quadrat-Test zeigten sich statistisch signifikante Unterschiede zwischen der ‚Syntaxgruppe' und der ‚Semantikgruppe'. Tendenzen, die sich in den Rohdaten der Teilnehmenden finden, lassen sich statistisch nicht weiter belegen. Für eine weitere Diskussion dieser Befunde vgl. Kapitel 5.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

5 Diskussion und Ausblick

Ein wichtiges Ziel dieser Arbeit war es, zweitsprachliche syntaktische Entwicklungsverläufe erwachsener Lernender detailliert zu modellieren und dabei Erkenntnisse darüber zu gewinnen, wann und in welchen gegenseitigen Abhängigkeiten sich ausgewählte syntaktische Strukturen entwickeln. Dies wurde auf der Ebene der Wortstellung untersucht, wobei zu diesem Zweck minimale satzwertige Einheiten als analytische Grundeinheit in dieser Studie operationalisiert wurden. Diesen wurden syntaktische Profil- bzw. Erwerbsstufen zugeordnet, die dann weitere Analysen ermöglichten. Zusätzlich wurden aus allen schriftlichen Texten drei syntaktische Komplexitätsmaße berechnet, bei denen davon ausgegangen wurde, dass sie stellvertretend für syntaktische Teilbereiche Aufschluss über Entwicklungsprozesse geben können.

Im Folgenden werden von den in dieser Arbeit bearbeiteten Fragestellungen ausgehend die zentralen Ergebnisse der vorliegenden Arbeit im Kontext bisheriger Forschung diskutiert. Ein Schwerpunkt liegt auf der Auseinandersetzung mit den methodischen Grundlagen und der Frage, inwiefern die gewählten Methoden geeignet waren, um zweitsprachliche Entwicklungen sichtbar und analysierbar zu machen. Ein Anliegen ist es zudem, auf methodischer Ebene Möglichkeiten herauszuarbeiten, wie sich die einbezogenen Perspektiven unter Umständen ergänzen könnten. In dieser Studie wird der Versuch unternommen, sich der schriftlichen zweitsprachlichen Entwicklung aus verschiedenen Perspektiven zu nähern und Erkenntnisse darüber zu gewinnen, was zu unterschiedlichen Zeitpunkten innerhalb dieser Entwicklung beobachtbar ist.

5.1 Heterogenität als Ausgangspunkt

Die in der vorliegenden Studie untersuchte Stichprobe von Lernerinnen und Lernern war darauf angelegt, durch die Kontrolle von ausgewählten Parametern eine größtmögliche Vergleichbarkeit bzw. ‚Homogenität' innerhalb der Stichprobe herzustellen. Hierzu wurden zahlreiche biografische und sprachliche Hintergründe der Teilnehmenden zur Teilnahme an der Studie vorausgesetzt (vgl. zu den einzelnen Voraussetzungen Kapitel 3.1.1). Zusätzlich wurden Unterricht und eingesetztes Material in allen Kursen mit Ausnahme der experimentellen Manipulation konstant gehalten. Die Lernenden besaßen einen hohen Grad an Vorbildung und lernten über anderthalb Jahre in einem Intensivkurs Deutsch als Zweitsprache und sind folglich nur sehr bedingt mit den Lernenden der Studien von Pienemann oder Clahsen aus den 1980er-Jahren vergleichbar. Vielmehr handelte es sich um Lernende, die in einem stark gesteuerten Kontext Deutsch als Zweitsprache lernten. Dazu kommt, dass die in der vorliegenden Studie untersuchten Lernerinnen und Lerner innerhalb der ausgewählten Teilstichprobe mehr Unterricht besuchten und daher mehr Input erhielten als die übrigen Lernenden innerhalb der Studie (vgl. Kapitel 3.2.1).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Es handelte sich bei den Lernerinnen und Lernern in dieser Studie folglich um eine sehr lernaffine Gruppe mit hoher Vorbildung. Dies spiegelte sich sicherlich in den Entwicklungsverläufen der Lernenden und muss bei der Interpretation der Ergebnisse und der herausgearbeiteten Entwicklungen mitbedacht werden. Grundsätzlich ist trotzdem sowohl anhand der Hintergrunddaten als auch der Sprachdaten festzustellen, dass in keinerlei Hinsicht von einer homogenen Gruppe von Lernenden ausgegangen werden kann und sich die zweitsprachlichen Entwicklungsverläufe der hier beobachteten Lernenden, in diesem Fall hauptsächlich Geflüchtete, durch sehr viele inter- und intraindividuelle Unterschiede auszeichnen. Dies zeigte sich sowohl auf der Ebene der durchgeführten Profilanalysen als auch in den longitudinalen Clusteranalysen. Deutlich wurde dieser Umstand außerdem an den Einzelfallanalysen der beiden Geschwister, die innerhalb eines kontrollierten Datensets zusätzlich für eine noch größere Ähnlichkeit zwischen den untersuchten Lernenden stehen (vgl. hierzu die nächsten beiden Kapitel der Diskussion).[226]

Obwohl das Spannungsfeld zwischen Homogenität und Heterogenität der gebildeten Stichprobe nicht explizit den Fokus dieser Untersuchung bildete, sollen an dieser Stelle mögliche Implikationen der vorliegenden Daten kritisch diskutiert werden. Heterogenität als Ausgangspunkt der Betrachtung zweitsprachlicher Erwerbsprozesse wird in jüngster Zeit in verschiedenen Beiträgen im Fachzusammenhang DaF/DaZ hervorgehoben (vgl. Gamper et al. 2020). Hünlich et al. (2018) gehen davon aus, dass die Heterogenität der Erwerbsumstände in der von den Autor/-innen analysierten Lernendengruppe sich hinsichtlich der „Lebensabschnitte und Erwerbsorte (Glaubensgemeinschaft, Schularten, Universität, Berufe, Transitländer, Auslandsaufenthalte etc.)“ (Hünlich et al. 2018: 24) als so groß darstelle, dass klare Unterscheidungen zwischen Erwerbskontexten weder möglich noch sinnvoll seien. Ähnlich stellt sich die Situation für die in dieser Studie untersuchte Stichprobe dar.

Die Annahme von homogenen L2-Lernendengruppen scheint nur sehr stark eingeschränkt möglich oder müsste unter Umständen komplett fallen gelassen werden. Die Beschreibung einer Gruppe aufgrund eines Charakteristikums (L1 oder DaZ-Lerner/-in) oder der Kombination von mehreren Charakteristika (L1, Alter, Bildungshintergrund etc.) ist kaum zielführend, da eben nur in manchen Aspekten eine ‚scheinbare‘ Homogenität innerhalb der untersuchten Gruppe erreicht werden kann. Gleichzeitig gerät die Annahme von sich homogen verhaltenden

226 Aus diesen grundsätzlichen Unterschieden zwischen den Lernerinnen und Lernern ergeben sich eine Reihe weiterer Forschungsperspektiven und Fragestellungen, die anhand der vorliegenden Daten untersucht werden könnten. Ein entscheidender Faktor, der bis jetzt nicht in die Analysen eingeflossen ist, wäre etwa ob sich Effekte aufgrund des Alters der Lernenden herausarbeiten lassen. Auf das Desiderat fehlender longitudinaler Studien, die sich mit dem Altersfaktor im Zweitsprachenerwerb beschäftigen, weist beispielsweise Czinglar (2014: 10) hin.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

L1-Sprechenden in jüngster Zeit zunehmend unter Druck. Vielmehr scheint es so zu sein, dass sich scheinbar ‚homogene' L1-Gruppen ebenfalls durch große interindividuelle und intraindividuelle Unterschiede auszeichnen (vgl. Shadrova et al. 2021: 21f.).[227] In diesem Kontext stellt sich dann die Frage, inwiefern eine L1-Norm in Bezug auf die Bewertung von „gelingendem Sprachhandeln" (Fandrych / Wallner 2022: 205) von L2-Lernenden sinnvoll sein kann. Es scheint sowohl für L1- als auch für L2-Sprechende unzählige und zum Teil stark differierende Möglichkeiten zu geben gelingendes Sprachhandeln zu realisieren.

Werden diese Diskussionen des Fachkontextes DaF/DaZ zusammengebracht, ergeben sich große methodische Herausforderungen bei der Analyse von Gruppendaten und damit die Notwendigkeit, bestimmte Annahmen und auf diesen beruhende didaktische Implikationen sehr vorsichtig zu formulieren. Diese Herausforderungen beziehen sich darauf, dass im Studiendesign und in der Auswertung von Daten unbedingt von Verallgemeinerungen, die auf einzelnen Faktoren beruhen, abgesehen werden sollte.[228] Gerade dann, wenn Gruppen von Lernenden in den Blick genommen werden, etwa arabischsprachige Deutschlernende, sollten Ergebnisse nicht weiter als notwendig verallgemeinert werden bzw. es muss sehr transparent dargestellt werden, was die untersuchte Gruppe tatsächlich gemeinsam hat und wo sich die untersuchten Individuen in ihren Voraussetzungen und in ihren Entwicklungsverläufen unterscheiden.[229] In der vorliegenden Studie diente die gemeinsame Erstsprache der Lernenden als zusätzlicher Kontrollparameter, der die angenommene Ähnlichkeit der Lernenden vergrößern sollte und gleichzeitig Möglichkeiten für weitere Forschungsarbeiten eröffnet, die Arabisch als Erstsprache systematisch in die Analyse einbeziehen könnten. Auch diese müssten dann aber aufgrund der Diglossie-Situation

227 Bereits Han (2004) macht darauf aufmerksam, dass die einzig mögliche exakte Definition einer L1-Sprecherin die ist, dass es sich bei ihr nicht um eine L2-Sprecherin handelt (vgl. Han 2004: 166), was verdeutlicht, wie unsicher die Definition sich darstellt.

228 Dass oft sehr weitreichende Verallgemeinerungen implizit reproduziert werden, verdeutlicht etwa der Band von Roche / Terrasi-Haufe (2016) zum Deutschunterricht mit Geflüchteten. Einerseits wird zwar auf die große Heterogenität der Zielgruppe verwiesen (vgl. Roche / Terrasi-Haufe 2016: 158), andererseits gehen die Autor/-innen davon aus, dass Lesen und Schreiben Geflüchteten generell schwerer falle als das Hören und Sprechen (vgl. ebd.: 134) oder dass der Umgang mit digitalen Lehr- und Lernmaterialien bei Geflüchteten nicht vorausgesetzt werden kann (vgl. ebd.: 156). Während dies sicherlich für viele Lernende, aber keinesfalls für ‚Geflüchtete' generell zutrifft, verdeutlichen Anmerkungen wie diese doch die Tendenz zu Verallgemeinerungen in sprachdidaktisch bezogenen Ansätzen.

229 Unbedingt sollten vor diesem Hintergrund methodische Ansätze kritisch hinterfragt werden, die beispielsweise ‚Fehlertypen' oder Ausspracheabweichungen monokausal auf (mögliche) Effekte oder Interferenzen einer Erstsprache beziehen.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

des Arabischen einerseits und der dialektalen Prägung der Umgangssprachen andererseits von grundsätzlich heterogenen Sprachhintergründen ausgehen.

Diese Diskussionen verdeutlichen in einem größeren Kontext, dass forschungsbezogene Ansätze im Zusammenhang von DaF/DaZ grundsätzlich von Heterogenität zwischen den untersuchten Lernerinnen und Lernern ausgehen sollten und sich dies zusätzlich in der Wahl geeigneter Methoden und dem transparenten Berichten aller relevanter Parameter zeigen müsste (vgl. Marsden / Plonsky 2018). Zuschreibungen von Lernenden als ‚DaZ-Lernende' besitzen letztlich nur bedingte Aussagekraft und müssten mit einer Reihe von weiteren Informationen ergänzt werden (vgl. Petersen 2019: 11). Dies ist vor allem dann von besonderer Relevanz, wenn es etwa um Fördermaßnahmen geht (vgl. ebd.). Es scheint in vielen Fällen nach wie vor ungeklärt, wie die Faktoren, die diese grundsätzliche Heterogenität der Lernerinnen und Lerner konstituieren, zweitsprachliche Erwerbsprozesse jeweils konkret beeinflussen und wie diese Faktoren im Zeitverlauf miteinander interagieren (vgl. Kuhs 2020: 531). Hier besteht ein großes Desiderat darin, genauer zu erforschen, wie „migrationsgesellschaftliche Neuformationen" (Knappik / Dirim 2018: 170) des Deutschen als Zweitsprache funktionieren und inwiefern diese mit lernersprachlichen Entwicklungen zusammengedacht und zusammengebracht werden können.

5.2 Methodische Diskussion und Limitationen der Studie

Methodisch versuchte die vorliegende Studie, einigen der im letzten Absatz besprochenen Herausforderungen dadurch entgegen zu treten, dass Verfahren aus dem CDST-Spektrum angewendet werden, die darauf abzielten, Variabilität in den Daten sichtbar zu machen. Diesen Verfahren liegt gleichzeitig die Annahme zugrunde, dass Lernen als Prozess untersucht werden muss (vgl. Lowie / Verspoor 2015: 67). Diese Studie stellte darüber hinaus einen Versuch dar, statistische und visuelle Analyseverfahren, die bis jetzt zum Teil in unterschiedlichen theoretischen Ansätzen operationalisiert werden, in ihrer Anwendung miteinander zu integrieren, um möglichst unterschiedliche Perspektiven auf lernersprachliche Daten zu ermöglichen.

Die Ergebnisse der Arbeit zeigen, dass die angewendeten Methoden und ihre Kombination relevante und neuartige Ergebnisse und eine sehr differenzierte Perspektive auf die zweitsprachliche Entwicklung von Lernenden geben und die Fragestellungen der vorliegenden Studie weitgehend beantwortet werden konnten. Alle verwendeten analytischen Verfahren haben spezifische Einschränkungen und können immer nur Teilbereiche der zweitsprachlichen Entwicklung fokussieren und dies sicherlich nicht erschöpfend. Während manche der Methoden wie die Profilanalyse als etabliert gelten können, kamen einige der verwendeten Methoden im Kontext DaF/DaZ und anhand deutschsprachiger Texte bisher m. W. zum ersten Mal zur Anwendung. Dies gilt vor allem für die CDST-inspirierten Methoden der longitudinalen Clusteranalysen, der glei-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

tenden Korrelationen und der *Change Point*-Analysen. Gerade die Integration unterschiedlicher Methoden könnte die Einschränkungen der einzelnen Verfahren in Teilen abschwächen.

Die in dieser Studie durchgeführten Profilanalysen nach Grießhaber lieferten die minimalen satzwertigen Einheiten, die dann in den verwendeten Komplexitätsmaßen die Grundlage der Operationalisierung bildeten. Dabei gingen in dieser Arbeit alle von den Lernenden produzierten Strukturen in die Analysen ein, ohne dass Strukturen ausgeschlossen wurden. Dies verdeutlicht eine wichtige Differenz zwischen den Ansätzen der *Processability Theory* und der Profilanalyse nach Grießhaber bei der Aufbereitung von Sprachdaten. In PT-basierten Studien wie Pienemann (1998) werden Strukturen, die im Verdacht stehen, Chunks zu sein, oft systematisch aus den Analysen ausgeschlossen. In der vorliegenden Studie wurde hingegen ein Ansatz gewählt, der sich strikt auf die tatsächlich von den Lernenden produzierten Strukturen bezieht und dabei mögliche Chunks oder feste Formeln in diesen Produktionen nicht aus den Analysen ausschließt. Dies führte dazu, dass unter Umständen deutlich mehr Strukturen als etwa in der Studie von Czinglar (2014) in die Analysen aufgenommen wurden. Ebenso wurden mögliche textuelle Übernahmen aus den Aufgabenstellungen nicht systematisch als Faktor in die Analysen einbezogen. Vielmehr wurde davon ausgegangen, dass alle produzierten Strukturen aus Sicht der Lernenden kommunikative Funktionen erfüllen und daher in die Analysen miteinbezogen werden sollten. Dazu kommt, dass ausschließlich anhand der schriftlichen Produktionen nicht bestimmbar ist, zu welchem Zeitpunkt Lernende Chunks für sich analytisch auflösen können und in ihren strukturellen Eigenschaften wahrnehmen. Eine Struktur wie *Wie geht's?*, die oft im Verdacht steht, ein Chunk zu sein, könnte zwar zu bestimmten Punkten der zweitsprachlichen Entwicklung von den Lernenden als im Ganzen gespeicherte Einheit produziert werden, aber zu anderen Zeitpunkten von den Lernenden in ihren Einzelteilen wahrgenommen und verstanden werden. In diesem Fall wäre eine Einstufung als Chunk unzulässig. Solche Strukturen wurden aufgrund dieser methodischen Unsicherheit in der vorliegenden Studie nicht aus den Analysen ausgeschlossen.

Aus den Profilanalysen ergaben sich die erreichten Profilstufen der Lernerinnen und Lerner zu jedem einzelnen Zeitpunkt, was eine wichtige Voraussetzung der longitudinalen Perspektive dieser Studie bildete. In den ersten sechs Monaten konnten die Ergebnisse der Profilanalyse auf den Erwerbsstufen 0–4 signifikante Unterschiede in der Entwicklung der Lernenden aufzeigen. Ab einem bestimmten Zeitpunkt lieferten die Ergebnisse der Profilanalyse auf Gruppenebene allerdings keine aussagekräftigen Ergebnisse mehr, da vom Großteil der Teilnehmenden die Profilstufe 4 ‚Verbendstellung' erreicht wurde. Solche Deckeneffekte, die die Trennschärfe der Profilanalyse beeinträchtigen, arbeiten ebenfalls Ehl et al. (2018) heraus. Dies ist nach etwa sechs Monaten, also zum Abschluss der ersten Studienphase, der Fall und deckt sich mit dem Zeitpunkt, an dem die meisten der Lernenden in dieser Studie die B1-Prüfung ablegten und

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

bestanden haben. Auffallend ist, dass Strukturen der beiden höheren Stufen 5 ‚Integrierter Nebensatz' und 6 ‚Erweitertes Partizipialattribut' nur sehr vereinzelt auftreten und keiner der Lernertexte zu einem der untersuchten Zeitpunkte auf einer dieser Niveaustufen eingestuft wurde. Dies gilt sowohl für die drei Zeitpunkte in den Gruppendaten als auch für die 24 untersuchten Zeitpunkte in den Einzelfallanalysen.[230] Die Profilanalyse scheint hinsichtlich der hier vorliegenden Ergebnisse nur für einen Teilzeitraum der zweitsprachlichen Entwicklung und nur für die in den Ansätzen der *Processability Theory* empirisch herausgearbeiteten Profilstufen relevante Ergebnisse zeigen zu können. Die von Grießhaber vorgeschlagenen, aber nicht empirisch belegten Profilstufen 5 und 6 spielen in der zweitsprachlichen Entwicklung der hier untersuchten Lernenden (noch) keine Rolle. Damit stimmen die Ergebnisse mit Analysen überein, die De Carlo / Gamper (2015) anhand der Daten von Schülerinnen und Schüler der Klassenstufen 2–6 herausarbeiten (vgl. De Carlo / Gamper 2015: 116). Dass die Profilstufen 5 und 6 im Gegensatz zu den Stufen 1–4 in den hier vorliegenden Daten kaum zu belegen sind, könnte damit zusammenhängen, dass es sich bei diesen höheren Stufen um Profilstufen handelt, die stark registerabhängig geprägt sind und daher eher einen Bezug zu bildungssprachlichen Kontexten aufweisen. Gerade in den letzten Monaten der Studie bereiteten sich die Lernenden allerdings auf die telc B2-Prüfung und die TestDaF-Prüfung vor und trotzdem können kaum Belege dieser beiden Stufen ausgemacht werden. Ein weiterer Grund für die eingeschränkte Aussagekraft der erreichten Profilstufen ist ihre grundsätzlich sehr breit angelegte Konzeption, durch die lernersprachliche Variation unter Umständen nicht sichtbar wird (vgl. Goschler 2017: 75). Es werden auf den einzelnen Stufen sehr unterschiedliche Strukturen zusammengefasst, die nicht einfach gleichgesetzt werden können. Dies gilt sowohl für die Profilstufen nach Grießhaber als auch für die Erwerbsstufen der *Processability Theory*. Unter der Stufe Inversion werden beispielsweise eine Reihe von Strukturen gefasst, die sich nicht unbedingt in den lernersprachlichen Daten der Teilnehmenden in dieser Studie wiederfinden. Während sich beispielsweise schon zu Beginn der Datenerhebung Inversionsstrukturen belegen lassen, bei denen adverbiale Phrasen ins Vorfeld gestellt wurden, finden sich bis zum Ende der Studie kaum Dativ- oder Akkusativobjekte oder Nebensätze im Vorfeld. Für die Stufe 4 ‚Verbendstellung' verhält sich die Situation ähnlich. Bestimmte Wortstellungsmuster werden hier frequenter und früher produziert als andere. Trotzdem gehen die stufenbezogenen Erwerbsmodelle davon aus, dass alle diese Strukturen prinzipiell zur Verfügung stehen, sobald eine bestimmte Erwerbsstufe erreicht wird. Die in dieser Studie ausgewerteten empirischen Daten stützen diese Annahme nicht, sondern lassen vielmehr den Schluss zu, dass bestimmte Strukturen (wie etwa Inversionen mit Objekten im Vorfeld), die theoretisch niedrigeren Profilstufen zugeordnet sind, den Lernenden selbst dann nicht zur

230 Tatsächlich wurde in allen schriftlichen Daten des gesamten MPI-HI-Korpus kein einziger Text diesen Profilstufen zugeordnet.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Verfügung stehen, wenn unterschiedliche Strukturen auf höheren Profilstufen produziert werden. Es müsste folglich innerhalb der einzelnen Profilstufen viel detailliertere und kleinteiligere Abstufungen geben, um dann zu überprüfen, inwiefern sich tatsächlich sequenzielle und implikationelle Entwicklungen rekonstruieren lassen. In diesem Kontext wären einerseits eine genauere Analyse einzelner Wortstellungsmuster, wie sie von Czinglar (2014) durchgeführt wird, und anderseits die gezielte Analyse der Variation bzw. der Variabilität innerhalb der angenommenen Stufen notwendig (vgl. Nicholas et al. 2019: 396–397).

Die Limitationen der Profilanalyse wurden deutlich sichtbar. Gerade auf einzelne Zeitpunkte bezogen können Profilanalysen ein letztlich irreführendes Bild auf den Entwicklungsstand von Lernenden geben. Dies wurde in den Daten der beiden Einzelfallanalysen und im Kontext der B1-Prüfungen deutlich, wenn die teilnehmende Person TN024 plötzlich eine deutlich niedrigere Profilstufe erreichte als in den Wochen und Monaten zuvor. Ehl et al. (2018) berichten für die von ihnen untersuchten Kinder ebenfalls, dass diese nach einigen Monaten niedrigere Profilstufen erreichen als zuvor (vgl. Ehl et al. 2018: 1269). Die Profilanalyse nach Grießhaber ist grundsätzlich als diagnostisches Instrument für den Einsatz zu einem Zeitpunkt konzipiert und soll anhand ihrer Ergebnisse verlässliche Aussagen zum Entwicklungsstand von Lernenden liefern. Dieser Anspruch erscheint vor dem Hintergrund dieser Arbeit fraglich. Die von Grießhaber (2012: 166–167) angeregte Übertragung der Methode auf weitere Bereiche der Sprachentwicklung muss in diesem Kontext ebenso kritisch gesehen werden. Es ist möglich, dass die Aussagekraft der Profilanalysen durch den spezifischen Kontext dieser Arbeit, die ausschließlich die schriftlichen Produktionen erwachsener Lernender untersucht, die zudem einen Intensivsprachkurs besuchen, zusätzlich eingeschränkt wird. Dennoch werfen die Ergebnisse dieser Arbeit die Frage auf, inwiefern Profilanalysen verlässlich sein können, falls sie nur einen Zeitpunkt fokussieren.

Die Clusteranalysen konnten auf zwei Ebenen wichtige Ergebnisse zeigen. Einerseits verdeutlichten sie zu jedem der drei Zeitpunkte allgemeine Entwicklungsschritte der Lernenden. Andererseits konnten sie in longitudinaler Perspektive Gruppen von Lernenden identifizieren, die sich ähnlich entwickeln. Diese Ähnlichkeit konnte unter anderem auf die Variabilität in den Daten der Komplexitätsmaße der Teilnehmenden bezogen werden, allerdings nicht für alle der untersuchten syntaktischen Komplexitätsmaße. Hinsichtlich der longitudinalen Clusteranalysen müssen vor allem zwei Aspekte kritisch angemerkt werden. Aufgrund der kleinen Stichprobengröße konnten den Analysen nur drei Clustervariablen zugrunde gelegt werden. Dies führte dazu, dass zwar gut begründete und im Rahmen dieser Studie auf ihre Eignung überprüfte Komplexitätsmaße eingesetzt wurden, diese aber zwangsläufig nur einen Bruchteil der syntaktischen Entwicklung der Lernenden abbilden konnten. Eine größere Stichprobengröße könnte hier durch den möglichen Einschluss weiterer Variablen Raum für komplexere Vergleichsdimensionen bilden, würde aber gleich-

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

zeitig zu einer Vielzahl weiterer Gruppen mit ähnlichen Entwicklungswegen bzw. zu einer größeren Anzahl an longitudinalen Clustern führen. Diese könnten dann wiederum die Interpretierbarkeit der Daten erheblich einschränken (vgl. Bulté / Housen 2020: 231).[231] Die Analysen zeigen, dass die drei Variablen nicht zu jedem Entwicklungszeitpunkt ähnlich viel zur Bildung distinkter Cluster beitragen (konnten). Unterschiede in der Clusterzusammensetzung ergaben sich vor allem aufgrund der Länge minimaler satzwertiger Einheiten und weniger stark aufgrund der Subordinationsrate oder der Rate komplexer Nominalphrasen.

Ein weiterer Punkt, der kritisch reflektiert werden muss, bezieht sich auf die Anzahl an Zeitpunkten, die in die longitudinalen Analysen auf Gruppenebene eingeflossen sind und auf die Frage, ob diese Messintervalle es zulassen, Aussagen über die Entwicklung der Lernenden zu treffen (vgl. de Bot et al. 2007; Baba / Nitta 2014: 28).[232] Einerseits war die Anzahl der Zeitpunkte ausreichend, um distinkte Gruppen von Lernenden mit jeweils ähnlichen Entwicklungsmustern in Bezug auf verschiedene Komplexitätsmaße zu ermitteln. Andererseits war die Form dieser Muster nur schwer zu interpretieren. Anhand von drei Zeitpunkten können kaum Rückschlüsse auf etwaige u-förmige Entwicklungen etc. gezogen werden (vgl. VanPatten 2014: 110). Zusätzlich ist daran zu erinnern, dass zu Zeitpunkt 3 die Daten aus der Probeprüfung und nicht aus der tatsächlichen Prüfung des Goethe-Zertifikats B1 in die Auswertung eingegangen sind. Es muss davon ausgegangen werden, dass die Motivation in einer Probeprüfung etwas niedriger als in einer offiziellen Prüfung ist, die gleichzeitig eine High-Stakes-Prüfung für die Lernerinnen und Lerner darstellt.

Vor allem die Variabilitäts- bzw. Trendanalysen anhand der Visualisierungen der Zeitreihendaten innerhalb der Einzelfallanalysen erwiesen sich bei der Beantwortung der Forschungsfragen als sehr hilfreiches Werkzeug. Sie visualisieren die sonst in den Daten verborgene Variabilität. Dies ermöglicht eine sehr feingranulare Modellierung zweitsprachlicher Entwicklungen, wie sie in der vorliegenden Studie unternommen wurde. Zusammen mit den *Change Point*-Analysen ermöglichten diese Verfahren es, Variabilität bestimmten Entwicklungen zuzuordnen (vgl. dazu den nächsten Punkt der Diskussion) und so Phasen zu erkennen, in denen es zu Neuordnungen des beobachteten syntaktischen Subsystems gekommen ist. Mithilfe der *Change Point*-Analysen konnten zudem

231 Möglich wäre hier sicherlich, zunächst mit möglichst vielen Variablen an das vorliegende Datenset heranzutreten und Variablen automatisiert berechnen zu lassen und dann statistisch herauszuarbeiten, welche Variablen oder Variablenkombinationen viel zur Varianz in den Daten beitragen (etwa durch eine *Principal Component Analysis*).

232 Lenzing (2015) allerdings erhebt in ihrer Studie nur zwei Datenpunkte, die durch ein Jahr getrennt sind. Trotzdem geht sie davon aus, dass im Kontext der *Processability Theory* gezeigt werden kann, dass Strukturen nicht übersprungen werden (Lenzing 2015: 110).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

statistisch signifikante Entwicklungssprünge in den Zeitreihendaten von TN024 und TN025 identifiziert werden. Vor allem die fehlenden *Change Points* in den Daten zu den erreichten Profilstufen überraschten hier allerdings ein wenig, da die Annahme einer Entwicklung über bestimmte Erwerbsstufen hinweg die Möglichkeit korrespondierender *Change Points* als wahrscheinlich erschienen ließ. Tatsächlich wurde in den Daten zu den erreichten Profilstufen für TN024 kein *Change Point* und für TN025 nur ein einzelner *Change Point* gefunden. In den *Change Point*-Analysen wurden außerdem Konfidenzintervalle herausgearbeitet, die sich zum Teil über mehrere Wochen oder sogar Monate innerhalb des Sprachkurses erstreckten. Die identifizierten Veränderungen müssen folglich mit Vorsicht interpretiert werden. Hier wäre in Zukunft zu klären, welche Aussagekraft und Bedeutung solche weit gefassten Konfidenzintervalle besitzen und wie sie besser eingegrenzt werden könnten.

Mithilfe der Variabilitäts- und Interaktionsanalysen konnten sehr effektiv Trends und Auffälligkeiten in den Zeitreihen identifiziert werden. Für die beiden Lernenden ergab sich eine Reihe unterschiedlicher Entwicklungsmuster. Die Analysen konnten beispielsweise sehr deutlich zeigen, dass dieselben Aufgabenstellungen bei den beiden Lernenden stark voneinander abweichende Textrealisierungen elizitierten. Zusätzlich konnten sich sehr einfach Zeitpunkte in den Zeitreihen identifizieren lassen, die deutlich von den anderen Werten abweichen. Vor allem die Interaktionsanalysen zu den Wortstellungsmustern gaben den Blick frei für Zusammenhänge, die anhand deskriptiver Statistiken nur schwer nachvollziehbar gewesen wären. Für TN024 konnte beispielsweise gezeigt werden, dass eine starke Präferenz für Inversionsstrukturen vorhanden ist, die sich über die gesamte Studiendauer zieht und Hinweise darauf gibt, dass TN024 schon zu Beginn seiner zweitsprachlichen Entwicklung komplexe textverknüpfende Strukturen einsetzen kann. Zudem demonstrierten die Interaktionsanalysen, dass TN024 im Gegensatz zu TN025 bei der Prüfung zum Goethe-Zertifikat B1 auf Strukturen niedrigerer Profilstufen zurückgreift als zu allen anderen Zeitpunkten. Sowohl für die Variabilitätsanalysen als auch für die Interaktionsanalysen bleiben noch Fragen darüber offen, ob sich statistisch belegen lässt, dass Zeitreihen der beiden sich hinsichtlich ihrer Variabilität unterscheiden und inwiefern solche Belege zielführend wären. Eine Möglichkeit dies weiter zu überprüfen wären etwa Monte Carlo-Analysen (vgl. Hou et al. 2020: 40). Diese wurden in der vorliegenden Studie nicht eingesetzt, sollten aber für zukünftige Analysen in die Überlegungen eingebunden werden.

Es liegt auf der Hand, dass die hier angewendeten Methoden für eine breitere Adaption im Bereich der Zweitspracherwerbsforschung genauer ausdifferenziert und in ihren Vor- und Nachteilen evaluiert und validiert werden müssten. Ein Punkt, der ebenfalls diskutiert werden muss, ist, inwiefern die von Vertreterinnen und Vertretern der CDST vorgeschlagenen und in dieser Studie angewendeten statistischen Verfahren mit den Annahmen der CDST übereinstimmen. Einer der Kritikpunkte, die aus der CDST vorgebracht werden, bezieht

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

sich auf das Zusammenfassen von individuellen Lernerdaten zu aggregierten Gruppendaten und die Annahme von letztlich linearen Entwicklungsverläufen. Die longitudinalen Clusteranalysen beispielsweise beruhen aber letztlich ebenfalls auf den Mittelwerten der einzelnen Cluster, die sich aus den aggregierten Gruppendaten ergeben. Die Anwendung von gleitenden Korrelationen, die auf das Abschwächen von linearen Trends zielt, setzt implizit voraus, dass lineare Trends in den Daten vorhanden sind (vgl. Bulté / Housen 2020: 230).

Wie die eingesetzten Verfahren haben die verwendeten Komplexitätsmaße Stärken und Schwächen. Längenmaße wie die in dieser Studie operationalisierte durchschnittliche Länge minimaler satzwertiger Einheiten müssen prinzipiell mit Vorsicht interpretiert werden, da es, ähnlich wie in erstsprachlichen Daten, dazu kommen kann, dass diese Komplexitätsmaße ein bestimmtes Plateau erreichen, auf dem die Lernenden sich nicht mehr unterscheiden und das gleichzeitig ein Deckenniveau für mögliche Entwicklung darstellt (Bulté / Housen 2012: 37). Komplexitätsmaße, bei denen es zu solchen Effekten kommt, sind dann nur in bestimmten Entwicklungsphasen hinsichtlich der zweitsprachlichen Entwicklung von Lernenden aussagekräftig. Hinsichtlich der anderen beiden Komplexitätsmaße (‚Subordinationsrate' und ‚Rate komplexer Nominalphrasen pro MSE') kann wie in anderen Studien (vgl. zum Beispiel Rosmawati 2020) beobachtet werden, dass subordinierende Variablen ab einem bestimmten Punkt zugunsten von steigenden Werten im Bereich der phrasalen Komplexität abnehmen, während die komplexen Nominalphrasen zunehmen und länger werden. Da allerdings in dieser Studie eine sehr enge Definition von komplexen bzw. erweiterten Nominalphrasen angelegt wird, sollte in Zukunft unbedingt überprüft werden, welchen Einfluss der Einbezug komplexerer Ausbaustufen auf die verwendeten Analysen hätte (vgl. Schellhardt / Schroeder 2015).

Eine weitere Eigenschaft der operationalisierten Komplexitätsmaße ist, dass sie zwar grundsätzlich Informationen über Entwicklungsprozesse liefern können, aber gleichzeitig keine Information darüber geben, wodurch diese Entwicklungen ausgelöst werden und was sie auf Textebene tatsächlich bedeuten (vgl. Rosmawati 2020: 114). Die kontextlose Zählung syntaktischer Strukturen kann nicht ausreichen, um das Konstrukt der syntaktischen Kompetenz klar zu umreißen (vgl. Ortega 2015: 91). Und ein höherer Grad an Komplexität in lernersprachlichen Texten „sagt noch nichts über ihre Angemessenheit und Funktionalität und somit über die pragmatische Dimension der Textqualität aus" (Petersen 2014: 80). Diese Perspektiven waren nicht Teil der vorliegenden Arbeit, aber es wird deutlich, dass hier weitere Studien nötig sind, die die textuelle Funktion bestimmter Strukturen, die zu bestimmten Zeitpunkten auftreten, genauer fassen und neben der lernersprachlichen Syntax weitere zweitsprachlichen Subsysteme in ihrer längsschnittlichen Entwicklung einbeziehen. Konkret auf das Schreiben bezogen bestehen für erwachsene Lernende Forschungslücken im Bereich der Entwicklung des schriftlichen Argumentierens (vgl. Domenech / Petersen 2018: 117).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

5.3 Variabilität und Erwerbsstufen

Inhaltlich war einer der Schwerpunkte dieser Arbeit die Frage, inwiefern sich für die untersuchte Stichprobe erwachsener Lernender mit Arabisch als Erstsprache Erwerbsstufen hinsichtlich der Entwicklung der Wortstellungsregeln des Deutschen rekonstruieren lassen. Forschungsfrage 1 dieser Studie bezog sich explizit auf die potenzielle Nachweisbarkeit von sequenziellen Erwerbsstufen. Die Existenz dieser Erwerbsstufen bzw. der in dieser Studie verwendeten Profilstufen zählt dabei zu den robustesten Annahmen über zweitsprachliche Entwicklung hinsichtlich des Deutschen als L2 (vgl. Ortega 2014: 176–177). Zahlreiche der Studien, die in dieser Arbeit besprochen wurden, bestätigen diese Sequenzen oder schlagen minimale Änderungen vor, die aber den grundsätzlich sequenziellen und implikationellen Charakter der Entwicklung im Bereich der Wortstellung nicht in Frage stellen. Die Analysen, die im Rahmen dieser Studie durchgeführt wurden, zeigen allerdings, dass der Nachweis von Erwerbssequenzen mit den gewählten Methoden nicht problemlos bzw. nur eingeschränkt möglich ist. Ein Grund für diesen Befund kann die Rolle der intraindividuellen Variabilität sein, die sich wie ein roter Faden durch die Argumentation dieser Arbeit zog und an zahlreichen Stellen sowohl auf Gruppenebene im Kontext der Clusteranalysen als auch in den Einzelfallanalysen herausgearbeitet wurde. Diese ist sicherlich multikausal und im Einzelnen nur schwer auf bestimmte Einzelfaktoren zurückzuführen. In der vorliegenden Studie wurde versucht, einige dieser vielfältigen Verbindungen in ihrer Komplexität sichtbar zu machen. Die CDST geht davon aus, dass Variabilität nicht nur charakteristisch für Lernersprache ist, sondern gleichzeitig wesentlich für die Entwicklung von Lernersprache verantwortlich ist (vgl. de Bot et al. 2007; vgl. Spoelman / Verspoor 2010: 532). Eine Schlussfolgerung dieser Perspektive ist, dass bei sich neu organisierenden Systemen vor und während diesen Veränderungen eine größere Variabilität zu beobachten ist als bei stabilen Systemen (vgl. Pinel / Csizér 2015: 185; vgl. Spoelman / Verspoor 2010: 536).

Die Ergebnisse des Friedmann-Tests, der testet, ob sich in der Gruppe der 44 Lernenden die erreichten Profilstufen im Zeitverlauf signifikant unterscheiden, legten zwar einen sequenziellen und weitgehend linearen Erwerbsverlauf, bei dem im Zeitverlauf immer höhere Profilstufen erreicht wurden, nahe, aber sie können dies nur auf Gruppenebene zeigen. Wurde für die 44 Teilnehmenden, deren Daten zu drei Zeitpunkten aus der ersten Studienphase untersucht werden, nur das erstmalige Auftreten einer Struktur gemäß eines angelegten Emergenzkriteriums herangezogen, konnten die vorausgesagten Sequenzen nur für einen kleinen Teil der Lernenden bestätigt werden. Die Emergenzanalysen zeigten allerdings ebenfalls, dass die hinsichtlich möglicher Erwerbssequenzen zunächst uneindeutigen Daten deutlichere Tendenzen zeigen, wenn eine breitere Datenbasis in die Beobachtung einbezogen wird. Wurden alle produzierten Strukturen zu allen drei Zeitpunkten in die Betrachtung miteinbezogen, konnten sich nur noch für drei der 44 Lernenden Belege dafür finden

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

lassen, dass ausschließlich höhere Profil- bzw. Erwerbsstufen erreicht werden, während zu allen drei Zeitpunkten Belege für niedrigere Stufen fehlen. Jansen (2008) macht in ihrer Studie darauf aufmerksam, dass sie nur für sieben von 21 Lernenden Belege dafür finden kann, dass bestimmte Stufen im Vergleich zu benachbarten Stufen noch nicht erworben sind (vgl. Jansen: 207). In den Daten von Lenzing (2015) zeigt sich ebenfalls, dass es zwar Lernende gibt, für die nicht ausreichend Belege existieren, die auf den Erwerb einer Stufe schließen lassen, aber gleichzeitig Belege dafür gefunden werden, dass eine höhere Erwerbsstufe bereits emergiert ist (ebd.: 111f.). Durch dieses Vorgehen wurde eine große methodische Herausforderung der Arbeit mit Lernersprachdaten im Kontext von Erwerbsprozessen deutlich. Wenn die Abwesenheit von Strukturen in einem Lernertext zu einem Zeitpunkt nicht gleichzeitig bedeutet, dass bestimmte Strukturen nicht als erworben gelten können, ist es im Grunde unmöglich den exakten Zeitpunkt bzw. den Startpunkt von Entwicklungen festzulegen. Dass Strukturen in unterschiedlichen Reihenfolgen emergent werden, muss daher nicht unbedingt heißen, dass sie auch in diesen erworben wurden. Es kann so letztlich nicht falsifiziert werden, dass Erwerbsstufen tatsächlich sequenziell und implikationell durchlaufen werden. Dies wiederum schränkt die Aussagekraft der verschiedenen Emergenzkriterien weiter ein, die außerdem nach wie vor einer empirischen Validierung entbehren. Vor allem die Einzelfallanalysen mit den Lernenden TN024 und TN205verdeutlichten aber im Gegenzug, wie schwierig es im Kontext dieser Arbeit war, die angenommenen Erwerbssequenzen in empirischen Daten wiederzufinden.

Die Emergenz der verschiedenen Profilstufen war vielmehr in den Daten nicht voneinander zu trennen. Die Stufen wurden von den beiden Lernenden zudem nicht nacheinander durchlaufen. Vielmehr traten schon zu Beginn der zweitsprachlichen Entwicklung Strukturen der Profilstufen 1,2 und 3 auf. Das heißt, dass bereits in den ersten Wochen der Schreibentwicklung fast alle der Stufen, die im Rahmen dieser Studie nachgewiesen werden konnten, eine Rolle spielen und von den Lernenden produktiv angewendet werden. Erwerbsstufen können per Definition nicht gleichzeitig erworben werden, weil niedrigere Stufen die Voraussetzung zur Verarbeitung der hierarchisch höheren Stufe darstellen (vgl. Jansen 2008) und daher als erworben gelten müssen, sobald höhere Stufen als erworben nachgewiesen werden. Sie können aber durchaus in engen zeitlichen Abständen emergieren (Pienemann 1998: 248). Die Messintervalle der vorliegenden Studie waren sehr eng gewählt und lieferten bereits nach wenigen Unterrichtseinheiten Daten der Teilnehmenden. Trotzdem konnten sich die einzelnen Profilstufen für die hier erhobenen Daten nicht auseinanderhalten lassen. Wenn trotz dieser Ergebnisse von der Gültigkeit der Erwerbssequenzen ausgegangen wird, bedeutet das gleichzeitig, dass die Profilanalyse als diagnostisches Instrument bereits nach sehr kurzer Zeit innerhalb der zweitsprachlichen Entwicklung ihre Trennschärfe verliert. Hier stellt sich die Frage, welche Rolle sequenzielle Erwerbsstufen spielen sollten, die kaum nachweisbar sind und welchen Mehrwert sie bringen, wenn darüber hinaus nicht klar ist, wie genau sich nachweisen lässt, dass die Stufen

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

tatsächlich noch nicht oder bereits erworben sind. Es konnten allerdings ebenfalls keine Belege für die Nicht-Existenz der Erwerbsstufen aus den Daten abgeleitet werden. Außerdem ist daran zu erinnern, dass die herausgearbeiteten Strukturen letztlich zu einem Teil auf subjektiven Präferenzen der Sprachverwendung der Teilnehmenden beruhen und es fraglich ist, ob diese ein objektives Bild des zugrundeliegenden zweitsprachlichen Systems oder der tatsächlich vorhandenen Kompetenzen liefern können (vgl. Bulté / Housen 2020: 213). Die im Unterricht erfasste Sprache deckt sich vielleicht nur teilweise mit der tatsächlich vorhandenen Sprache (vgl. Dirim / Krehut 2020).

Die Kombination der Variabilitätsanalysen und der *Change Point*-Analysen zeigte, dass jede in dieser Studie ermittelte statistisch signifikante Veränderung innerhalb der Zeitreihendaten der Einzelfallanalysen in eine Phase erhöhter Variabilität fällt. Dies deckt sich mit der Definition eines Phasenübergangs nach Baba / Nitta (2014), die eine große Fluktuation der beobachteten Variable als Voraussetzung für einen solchen Übergang sehen (vgl. Baba / Nitta 2014: 27). Gleichzeitig wurde aber nicht in jeder Phase großer Variabilität ein CP identifiziert. Bereits Meisel et al. (1981) beobachten Phasen großer Variabilität in ihren Daten, wobei sie ansteigende Variabilität mit dem Wechsel zwischen Entwicklungsstufen assoziieren und dies dann problematisieren (Meisel et al. 1981: 111). Gerade in Bezug auf die erreichten Profilstufen fanden sich in den Daten der Einzelfallanalysen der vorliegenden Studie nur wenige *Change Points*. Insgesamt scheint die Variabilität in diesen Datenreihen zu groß, um Entwicklungsschritte eindeutig statistisch identifizieren zu können. Bulté / Housen (2020) merken an, dass in vielen CDST-basierten Studien zwar ein Zusammenhang zwischen Variabilität und Entwicklung angenommen, aber nur selten belegt werden kann (vgl. Bulté / Housen 2020: 230). Die in dieser Studie präsentierten Daten legen einen deutlichen Zusammenhang zwischen beiden Konzepten nahe und zeigen darüber hinaus, dass Variabilität in lernersprachlichen Daten nicht zufällig auftritt und nicht einfach als ‚Hintergrundgeräusch' (vgl. Verspoor et al. 2008) betrachtet werden kann, sondern genau in den Phasen auftritt, in denen dann statistisch signifikante Veränderungen nachgewiesen werden können. Eine große Variabilität ist indikativ für eine potenzielle (signifikante) Veränderung innerhalb der zweitsprachlichen Entwicklung der beobachteten Lernenden. Folgestudien könnten beispielsweise untersuchen, ob sich Unterschiede zwischen Phasen großer Variabilität ohne *Change Point* und Phasen großer Variabilität mit *Change Point* ausmachen lassen.

Ein entscheidendes Ergebnis, das sich in allen der durchgeführten Analysen sowohl auf Gruppenebene als auch in den Emergenzanalysen und den Einzelfallanalysen zeigte, ist die starke Abhängigkeit der angelegten Analysen von einer geeigneten Datenbasis.

Ein letzter Punkt, der hinsichtlich der Ergebnisse dieser Studie diskutiert werden soll, ist die Annahme, dass die durch die PT und die PA vorausgesagten Erwerbssequenzen universelle Gültigkeit besitzen, d.h., dass alle Lernerinnen

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

und Lernern diesen Sequenzen folgen. Grundsätzlich müsste an dieser Stelle die Frage gestellt werden, wie viele der Lernerinnen und Lerner in ihrer Entwicklung einer bestimmten Sequenz folgen müssten, damit diese als universell bezeichnet werden könnte. Um diese Frage für das Deutsche als L2 zufriedenstellend beantworten zu können, wären einerseits Replikationsstudien mit sehr vergleichbaren Gruppen von Lernenden und unter Einsatz vergleichbarer methodischer Verfahren notwendig. Andererseits wäre eine große Datenbasis lernersprachlicher Produkte wie im englischsprachigen Lernerkorpus *EFcamdat* zielführend. Eine solch umfassende Datenbasis existiert für das Deutsche als L2 bis heute nicht einmal ansatzweise. Eine Möglichkeit, dies methodisch zu umgehen, wären in diesem Zusammenhang korpusübergreifende Analysen, die mehrere Lernerkorpora des Deutschen gleichzeitig einbeziehen.

5.4 Experimentelle Manipulation

Hinsichtlich der experimentellen Manipulation innerhalb der Studie ist für sprachdidaktische Kontexte die Frage relevant, ob grammatisch-syntaktische Inhalte explizit und extensiv im Unterricht thematisiert oder eher impliziert und beiläufig eingeführt und behandelt werden sollten (vgl. Fandrych 2013: 165). Die inferenzstatistischen Verfahren wurden im Rahmen dieser Arbeit allerdings nicht eingesetzt, um zu überprüfen, ob die experimentelle Manipulation ‚ein Erfolg war', d. h. ob sich tatsächlich Entwicklungsvorteile für eine der beiden Instruktionsgruppen zeigen. Vielmehr sollten Belege dafür gesammelt werden, ob potenzielle Effekte der experimentellen Manipulation die Ergebnisse der durchgeführten Methoden verzerren könnten.

Für die Berechnung von Gruppenunterschieden in Bezug die experimentelle Manipulation konnten sowohl anhand der ANOVAs als auch durch den Chi Square-Test keine statistisch signifikanten Unterschiede zwischen den Syntax- und Semantikgruppen herausgearbeitet werden. Für die in dieser Studie untersuchten Fragestellungen wird daraus der Schluss gezogen, dass durch die experimentelle Manipulation die Ergebnisse der übrigen Analysen nicht in einem Maße verzerrt wurden, die die Gültigkeit dieser Ergebnisse erheblich einschränkt. In weiteren Schritten könnte allerdings evaluiert werden, wie sich andere externe Variablen wie beispielsweise das Alter der Teilnehmenden, deren Vorbildung oder die Ergebnisse der Sprachtests zu den ermittelten longitudinalen Clustern verhalten (vgl. Pinel / Csizér 2015). Dies könnte auf einer höheren Ebene Aufschluss darüber geben, inwiefern externe Faktoren unter Umständen mit der beobachteten inter- und intraindividuellen Variabilität in Verbindung gebracht werden können. In diesem Zusammenhang könnten außerdem weitere Komplexitätsmaße in die Analysen einbezogen werden, die die begrenzte Auswahl in der vorliegenden Studie gezielt ergänzen und so weitere Informationen über die Wirksamkeit der durchgeführten Manipulation liefern könnten (vgl. Bulté / Housen 2012: 34).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Es kann mehrere Gründe dafür geben, dass keine Unterschiede bzw. Effekte der Manipulation gefunden wurden, die im Folgenden kurz angerissen werden sollen. Einige dieser Gründe haben mit der Quantität bzw. der Qualität der experimentellen Manipulation selbst zu tun. Einige haben einen eher methodischeren Schwerpunkt. Zunächst kann durchaus davon ausgegangen werden, dass der zeitliche Umfang der experimentellen Manipulation zu gering angesetzt war, um tatsächlich unterschiedliche Entwicklungen der beiden Instruktionsgruppen zu evozieren bzw., dass die gewählten Messintervalle unpassend waren, um Veränderungen sichtbar zu machen. In diesem Zusammenhang müsste darüber hinaus die Rolle der unterschiedlichen Lehrkräfte genauer in den Blick genommen werden, um auszuschließen, dass diese unter Umständen ebenfalls Effekte auf die unterschiedlichen Lerngruppen hatten. Ebenso ist es möglich, dass die Inhalte der experimentellen Manipulation letztlich unpassend gewählt wurden oder sich zu stark ähnelten, um Unterschiede zwischen den beiden Gruppen zu bedingen, da das übrige Curriculum der Lerngruppen identisch war. Als Letztes könnte es sein, dass die Lernenden durch die Lernsituation in Deutschland zusätzlich so viel weiteren und sich durch innersprachliche Variation auszeichnenden Input erhielten, dass mögliche Effekte der Manipulation der Unterrichtsgruppen nicht wirksam wurden (vgl. zum möglichen Einfluss innersprachlicher Variation auf DaZ-Lernende Lenzhofer-Glantschnig / Scherr 2015: 173).

Eine weitere, im Rahmen der vorliegenden Untersuchung ebenfalls nicht überprüfte Möglichkeit ist die, dass die Variabilität zwischen den Lernenden unter Umständen die Effekte der experimentellen Manipulation maskiert. In diesem Fall wäre die Hypothese, dass die unterrichtliche Intervention Wirkung gezeigt hätte, diese Wirkung aber durch die hohe interindividuelle Variabilität zwischen den Lernenden nicht sichtbar wird bzw. nachweisbar ist. In diesem Zusammenhang wäre beispielsweise möglich, dass Gruppenunterschiede zu Beginn der Studie existieren, die nicht bemerkt wurden. Da die Zuteilung zu den Experimentalgruppen zufällig erfolgte, ist diese Möglichkeit zwar unwahrscheinlich, jedoch nicht komplett ausgeschlossen. Es ist zuletzt durchaus möglich, dass sich Ergebnisse erst dann zeigen, wenn eine größere Anzahl von Variablen miteinbezogen wird, etwa semantische Variablen, mündliche Daten oder Daten, die sich auf rezeptive Tests beziehen. Ein bestimmter Input zu einer bestimmten Zeit muss nicht notwendigerweise einen bestimmten Output hervorrufen (vgl. VanPatten 2014: 107; Lowie / Verspoor 2015: 70).

Zusammenfassend wird an dieser Stelle deutlich, dass es hier eine Anzahl von Desiderata gibt, die in weiteren Untersuchungen und mit anderen als den hier gewählten methodischen Verfahren zu behandeln wäre. Neben den erwähnten Unsicherheiten, die in weitere Studien systematisch einbezogen werden müssten, wäre zu klären, ob es spezifische Variablen oder andere Komplexitätsmaße gibt, in denen sich Differenzen zeigen, die auf die experimentelle Manipulation zurückgeführt werden könnten und ob sich zwischen den Lerngruppen Unterschiede hinsichtlich der Lerngeschwindigkeit ergeben.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

5.5 Ausblick

Abschließend ergeben sich anhand der vorliegenden Ergebnisse mehrere konkrete Perspektiven für zukünftige Forschungsarbeiten mit den in dieser Studie vorgestellten Daten. Diese Perspektiven setzen die weitere Aufbereitung des zugrundeliegenden MPI-HI-Korpus und die Integration der bis jetzt noch nicht einbezogenen Daten voraus. Es wäre außerdem zu überlegen, wie die Ergebnisse der vorliegenden Studie in Beziehung zu den neurowissenschaftlichen Daten, die in der Gesamtstudie erhoben wurden, gesetzt werden könnten.

Ein dringendes Desiderat im Zusammenhang mit den Daten aus dem MPI-HI-Korpus wäre die weitere systematische Aufbereitung und die Veröffentlichung der schriftlichen sowie der mündlichen Daten, die im Rahmen der Studie gesammelt wurden. Dies konnte im Rahmen der Arbeit an dieser Dissertation nicht geleistet werden, bleibt aber eine attraktive und relevante Zukunftsaufgabe. Dies wäre vor allem deshalb vielversprechend, weil durch die durchgeführten Sprachtests innerhalb der Studie empirisch begründete Verbindungen zwischen GER-Niveaustufen und zweitsprachlicher Entwicklung gezogen werden könnten, die bis heute für das Deutsche nur vereinzelt vorliegen.

Ebenfalls wäre der systematische Einbezug mündlicher und rezeptiver Daten, die für die Lernenden vorliegen, in zukünftigen Studien von großer Relevanz, um die zweitsprachliche Entwicklung der hier beobachteten Lernenden kleinteiliger und umfassender aufzeigen zu können und verschiedene Modalitäten miteinander in Beziehung zu setzen. Dies könnte dabei helfen, mit den Daten derselben Lernenden herauszuarbeiten, wie sich verschiedene lernersprachlichen Systeme gleichzeitig (und unter Umständen unterschiedlich) entwickeln und gegenseitig beeinflussen. In dieser Arbeit wurde versucht, diese Interaktionen anhand weniger ausgewählter Variablen darzustellen. Dies hat notwendigerweise und gerade vor dem Hintergrund der Vielzahl an mittlerweile zur Verfügung stehenden Komplexitätsmaße eine starke Vereinfachung bzw. Reduktion der tatsächlich zugrundeliegenden Komplexität zur Folge. Aus diesem Grund wäre der Einbezug weiterer Parameter und mündlicher lernersprachlicher Daten vielversprechend. Nicht zuletzt wurden in der hier vorliegenden Arbeit syntaktische Aspekte des lernersprachlichen Systems betrachtet und hinsichtlich ihrer möglichen Interaktionen untersucht. Eine Erweiterung des Fokus der Untersuchung auf die Ebenen des Wortschatzes bzw. der semantischen Entwicklung oder auch der orthografischen Entwicklung wäre dringend geboten. Hier wären mit den vorgestellten Methoden etwa Studien zum Gebrauch bestimmter Wortarten oder der Entwicklung von Kollokationen und Mehrworteinheiten denkbar. Eine zusätzliche Perspektive bietet die Analyse der Entwicklung spezifischer syntaktischer Strukturen wie beispielsweise des Vorfeldes. Gemessen an der syntaktischen Relevanz des Vorfeldes und den Möglichkeiten, die Lernende durch den Erwerb der Vorfeldposition kommunikativ erlangen, stellt sich die Frage, inwiefern der Erwerb des Vorfeldes mit dem Erwerb anderer syntaktischer Wortstellungsregeln des Deutschen korreliert. Nicht zuletzt dient

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

das Vorfeld im Deutschen „zur textuellen Einbindung eines Satzes, zu dem es gehört" (Fandrych 2003: 178). Es ermöglicht den Lernenden folglich, Aussagen zu treffen, die über die Satzgrenze hinausgehen. Aus sprachdidaktischer Sicht könnte eine sichere Beherrschung des Vorfeldes und der zugrundeliegenden syntaktischen Regeln den Lernenden dabei helfen, in komplexen Satzgefügen nicht die Übersicht zu verlieren (vgl. ebd.). Im Zusammenhang mit der Entwicklung des Vorfeldes wären eine Reihe weiterführender Fragen verbunden. Wie teilen Lernende relevante Informationen in ihren schriftlichen Produktionen mit? Mit welchen Mitteln transportieren Lernende Informationen, wenn die in den jeweils obligatorischen Kontexten benötigten syntaktischen Strukturen nicht verwendet werden oder noch nicht zur Verfügung stehen?

Die Daten aus dieser Studie wurden im Rahmen einer Studie zur neuronalen Entwicklung während des Zweitspracherwerbs erwachsener Lernender gesammelt. Daraus ergibt sich in Bezug auf die hier vorliegenden Daten zuletzt das Desiderat, diese mit den neurowissenschaftlichen Daten derselben Lernenden zusammenzubringen. Hier wäre unter anderem die Frage relevant, inwieweit etwa die in den Clusteranalysen herausgearbeiteten Entwicklungsprofile in Zusammenhang mit der erforschten Plastizität neuronaler Netzwerke gebracht werden könnten und ob die Variabilität der zweitsprachlichen Entwicklungsverläufe in diesem Kontext operationalisiert werden könnte. Dieses letzte Desiderat verdeutlicht im Zusammenspiel mit den Ergebnissen dieser Studie die Notwendigkeit zukünftige Forschungsperspektiven im Kontext von DaF/DaZ noch stärker interdisziplinär zu denken und auszurichten. Besonders vielversprechende Perspektiven zur Weiterentwicklung der Analyse von Lernersprache(n) bieten hier die Computerlinguistik und die in den letzten Jahren stark weiterentwickelten sprachtechnologischen Ansätze.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

6 Literaturverzeichnis

Abel, Andrea / Wisniewski, Katrin / Nicolas, Lionel / Boyd, Adriane / Hana, Jirka / Meurers, Detmar (2014): A Trilingual Learner Corpus illustrating European Reference Levels. In: Ricognizioni – Rivista di Lingue, Letterature e Culture Moderne, 2/1, 111–126.

Abutalebi, Jubin / Green, David (2007): Bilingual language production: the neurocognition of language representation and control. In: Journal of Neurolinguistics, 20/3, 242–275. DOI: 10.1016/j.jneuroling.2006.10.003.[233]

Ahrenholz, Bernt (Hg.) (2012): Einblicke in die Zweitspracherwerbsforschung und ihre methodischen Verfahren. Berlin / Boston: De Gruyter. DOI: doi.org/10.1515/9783110267822.

Ahrenholz, Bernt (2012a): Methodische Verfahren der Zweitspracherwerbsforschung – zur Einführung. In: Ahrenholz, Bernt (Hg.): Einblicke in die Zweitspracherwerbsforschung und ihre methodischen Verfahren. Berlin / Boston: De Gruyter, 1–26. DOI: 10.1515/9783110267822.1.

Ahrenholz, Bernt (2014): Lernersprachenanalyse. In: Settinieri, Julia / Demirkaya, Sevilen / Feldmeier, Alexis / Gültekin-Karakoç, Nazan / Riemer, Claudia (Hg.): Empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache. Eine Einführung. Paderborn: Ferdinand Schöningh, 167–179.

Ahrenholz, Bernt (2020a): Erstsprache – Zweitsprache – Fremdsprache – Mehrsprachigkeit. In: Ahrenholz, Bernt / Oomen-Welke, Ingelore / Ulrich, Winfried (Hg.): Deutsch als Zweitsprache. 5. unveränderte Auflage. Baltmannsweiler: Schneider Hohengehren, 3–20.

Ahrenholz, Bernt (2020b): Zweitspracherwerbsforschung. In: Ahrenholz, Bernt / Oomen-Welke, Ingelore / Ulrich, Winfried (Hg.): Deutsch als Zweitsprache. 5. unveränderte Auflage. Baltmannsweiler: Schneider Hohengehren, 102–120.

Ahrenholz, Bernt / Oomen-Welke, Ingelore / Ulrich, Winfried (Hg.) (2020): Deutsch als Zweitsprache. 5. unveränderte Auflage. Baltmannsweiler: Schneider Hohengehren.

Albers, Sönke / Klapper, Daniel / Konradt, Udo / Walter, Achim / Wolf, Joachim (Hg.) (2009): Methodik der empirischen Forschung. Wiesbaden: Gabler.

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, 67/1, 180–208. DOI: 10.1111/lang.12232.

233 Wo immer möglich sind DOI (Digital Object Identifier) angegeben. Falls das nicht möglich ist, sind die Internetadresse und der Zeitpunkt des letzten Zugriffs in eckigen Klammern angegeben.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Altmayer, Claus (2022): Deutsch als Fremdsprache vs. Deutsch als Zweitsprache: Überlegungen zu einer bislang ausgebliebenen, aber notwendigen Debatte. In: Deutsch als Fremdsprache, 59/1, 15–26.

Altmayer, Claus / Biebighäuser, Katrin / Haberzettl, Stefanie / Heine, Antje (Hg.) (2021): Handbuch Deutsch als Fremd- und Zweitsprache. Stuttgart: J.B. Metzler. DOI: 10.1007/978–3–476–04858–5.

Anstatt, Tanja (Hg.) (2007): Mehrsprachigkeit bei Kindern und Erwachsenen. Erwerb, Formen, Förderung. Tübingen: Narr Francke Attempto.

Anwander, Alfred / Friederici, Angela D. (2015): Ethikantrag vom 4.12.2015 an die Ethik-Kommission der Universität Leipzig zur Studie: Strukturelle Plastizität des Sprachnetzwerkes beim Zweitspracherwerb von Erwachsenen.

Baba, Kyoko / Nitta, Ryo (2014): Phase Transitions in Development of Writing Fluency From a Complex Dynamic Systems Perspective. In: Language Learning, 64/1, 1–35. DOI: 10.1111/lang.12033.

Backhaus, Klaus / Erichson, Bernd / Plinke, Wulff / Weiber, Rolf (2018): Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. Berlin / Heidelberg: Springer Gabler. DOI: 10.1007/978–3–662–56655–8.

Baddeley, Allan (2003): Working memory and language: An overview. In: Journal of Communication Disorders, 36/3, 189–208. DOI: 10.1016/S0021–9924(03)00019–4.

Bailey, Nathalie / Madden, Carolyn / Krashen, Stephen D. (1974): Is there a "natural sequence" in adult second language learning? In: Language Learning, 24/2, 235–243. DOI: 10.1111/j.1467–1770.1974.tb00505.x.

Ballestracci, Sabrina (2008): Überindividuelle Merkmale des Grammatikerwerbs im Unterricht des Deutschen als Fremdsprache durch italophone Studierende. Ergebnisse einer empirischen Untersuchung. In: Deutsch als Fremdsprache, 45/3, 160–169. DOI: 10.37307/j.2198–2430.2008.03.06.

Baraldi, Amanda N. / Enders, Craig K. (2010): An introduction to modern missing data analyses. In: Journal of school psychology, 48/1, 5–37. DOI: 10.1016/j.jsp.2009.10.001.

Bardovi-Harlig, Kathleen (2014): Documenting interlanguage development. In: Han, ZhaoHong / Tarone, Elaine (Hg.): Interlanguage: Forty Years Later. Amsterdam: John Benjamins, 127–146. DOI: 10.1075/lllt.39.08ch6.

Bardovi-Harlig, Kathleen (2015): One functional Approach to SLA. The Concept-Oriented Approach. In: VanPatten, Bill / Williams, Jessica (Hg.): Theories in second language acquisition. An introduction. 2. Auflage. New York / London: Routledge, 54–74.

Barkaoui, Khaled (2014): Quantitative Approaches for Analyzing Longitudinal Data in Second Language Research. In: Annual Review of Applied Linguistics 34, 65–101. DOI: 10.1017/S0267190514000105.

Bassano, Dominique / van Geert, Paul (2007): Modeling continuity and discontinuity in utterance length: a quantitative approach to changes, transitions and intra-individual variability in early grammatical development. In: Developmental Science, 10/5, 588–612. DOI: 10.1111/j.1467–7687.2006.00629.x.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Baten, Kristof (2010): Die Erwerbssequenzhypothese: Theorie und Praxis des Kasuserwerbs. In: Deutsche Sprache, 38/1, 43–69. DOI: 10.37307/j.1868–775X.2010.01.04.

Baten, Kristof (2011): Processability theory and German case acquisition. In: Language Learning, 61/2, 455–505. DOI: 10.1111/j.1467–9922.2010.00615.x.

Baten, Kristof (2013): Der Genitiv als Attributskasus im Fremdsprachenerwerb. In: Deutsch als Fremdsprache, 50/4, 195–205. DOI: 10.37307/j.2198–2430.2013.04.03.

Baten, Kristof / Lochtman, Katja (2014): Das deutsche Kasussystem im Fremdsprachenerwerb: Ein Forschungsüberblick. In: Muttersprache, 124/1, 1–25.

Bausch, Karl-Richard (Hg.) (2004): Mehrsprachigkeit im Fokus. Arbeitspapiere der 24. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr.

Becker, Tabea (2018): Schreibentwicklung in der Grundschule. In: Grießhaber, Wilhelm / Schmölzer-Eibinger, Sabine / Roll, Heike / Schramm, Karen (Hg.): Schreiben in der Zweitsprache Deutsch. Berlin / Boston: De Gruyter, 79–93. DOI: 10.1515/9783110354577–006.

Becker-Mrotzek, Michael / Grabowski, Joachim / Steinhoff, Torsten (Hg.) (2017): Forschungshandbuch empirische Schreibdidaktik. Münster: Waxmann.

Bell, Melanie L. / Kenward, Michael G. / Fairclough, Diane L. / Horton, Nicholas J. (2013): Differential dropout and bias in randomised controlled trials: when it matters and when it may not. In: BMJ [British Medical Journal]. DOI: 10.1136/bmj.e8668.

Belz, Maria / Belz, Michael / Özkan, Ibrahim / Graef-Callies, Iris T. (2017): Posttraumatic stress disorder and comorbid depression among refugees: Assessment of a sample from a German refugee reception center. In: Transcultural Psychiatry, 54/5–6, 595–610. DOI: 10.1177/1363461517745473.

Berwick, Robert / Friederici, Angela D. / Chomsky, Noam / Bolhuis, Johan (2013): Evolution, brain, and the nature of language. In: Trends in Cognitive Sciences, 17/2, 89–98. DOI: 10.1016/j.tics.2012.12.002.

Bley-Vroman, Robert (1983): The comparative fallacy in interlanguage studies: the case of systematicity. In: Language Learning, 33/1, 1–17. DOI: 10.1111/j.1467–1770.1983.tb00983.x.

Bohnacker, Ute (2006): When Swedes begin to learn German: from V2 to V2. In: Second Language Research, 22/4, 443–486. DOI: 10.1191/0267658306sr275oa.

Bordag, Denisa / Opitz, Andreas (2021): Zweitspracherwerbsforschung. In: Altmayer, Claus / Biebighäuser, Katrin / Haberzettl, Stefanie / Heine, Antje (Hg.): Handbuch Deutsch als Fremd- und Zweitsprache. Stuttgart: J.B. Metzler, 192–212. DOI: 10.1007/978–3–476–04858–5_13.

Boss, Bettina (2004): Wann ich habe Freizeit, ich koche gern. Zum Erwerb der deutschen Inversion und Nebensatzwortstellung durch australische Studierende. In: Deutsch als Fremdsprache, 41/1, 28–32. DOI: 10.37307/j.2198–2430.2004.01.06.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Bot, Kees de (2017): Complexity Theory and Dynamic Systems Theory: Same or different? In: Ortega, Lourdes / Han, ZhaoHong (Hg.): Complexity Theory and Language Development. In Celebration of Diane Larsen-Freeman. Amsterdam / Philadelphia: John Benjamins, 51–58. DOI: 10.1075/lllt.48.03deb.

Bot, Kees de / Lowie, Wander / Verspoor, Marjolijn (2007): A Dynamic Systems Theory approach to second language acquisition. In: Bilingualism, 10/1, 7–21. DOI: 10.1017/S1366728906002732.

Bot, Kees de / Larsen-Freeman, Diane (2011): Researching Second Language Development from a Dynamic Systems Theory perspective. In: Verspoor, Marjolijn / de Bot, Kees / Lowie, Wander (Hg.): A dynamic approach to second language development. Methods and techniques. Amsterdam / Philadelphia: John Benjamins, 5–23. DOI: 10.1075/lllt.29.01deb.

Börner, Wolfgang / Vogel, Klaus (Hg.) (2002): Grammatik und Fremdsprachenerwerb. Kognitive, psycholinguistische und erwerbstheoretische Perspektiven. Tübingen: Narr Francke Attempto.

Braun-Podeschwa, Julia / Habersack, Charlotte / Pude, Angela (2015): Menschen B1. Kursbuch. München: Hueber.

Breitsameter, Anna / Glas-Peters, Sabine / Pude, Angela (2013): Menschen A2. Arbeitsbuch. Ismaning: Hueber.

Breitsameter, Anna / Glas-Peters, Sabine / Pude, Angela (2015): Menschen B1. Arbeitsbuch. München: Hueber.

Brezina, Vaclav (2018): Statistics in corpus linguistics. A practical guide. Cambridge: Cambridge University Press. DOI: 10.1017/9781316410899.

Brezina, Vaclav / Pallotti, Gabriele (2019): Morphological complexity in written L2 texts. In: Second Language Research, 35/1, 99–119. DOI: 10.1177/0267658316643125.

Brünger-Weilandt, Sabine / Bruhn, Kai-Christian / Busch, Alexandra W. / Hinrichs, Erhard / Horstmann, Wolfram / Grötschel, Martin / Paulmann, Johannes / von Rummel, Philipp / Schlotheuber, Eva / Schmidt, Dörte / Schrade, Torsten / Simon, Holger (2019): Memorandum of Understanding by NFDI initiatives from the humanities and cultural studies. DOI: 10.5281/zenodo.3265763.

Bryant, Doreen / Rinker, Tanja (2021): Der Erwerb des Deutschen im Kontext von Mehrsprachigkeit. Tübingen: Narr Francke Attempto.

Bulté, Bram / Housen, Alex (2012): Defining and operationalising L2 complexity. In: Housen, Alex / Kuiken, Folkert / Vedder, Ineke (Hg.): Dimensions of L2 Performance and Proficiency. Amsterdam: John Benjamins, 21–46. DOI: 10.1075/lllt.32.02bul.

Bulté, Bram / Housen, Alex (2020): A critical appraisal of the CDST approach to investigating linguistic complexity in L2 writing development. In: Fogal, Gary / Verspoor, Marjolijn (Hg.): Complex Dynamic Systems Theory and L2 Writing Development. Amsterdam: John Benjamins, 207–238. DOI: 10.1075/lllt.54.09bul.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Bundesamt für Migration und Flüchtlinge [BAMF] / Europäisches Migrationsnetzwerk [EMN] / Hoffmeyer-Zlotnik, Paula / Grote, Janne / Tangermann, Julian / Konar, Özlem (2016): Migration, Integration, Asyl. Politische Entwicklungen in Deutschland 2016. Jährlicher Bericht der deutschen nationalen Kontaktstelle für das europäische Migrationsnetzwerk (EMN). Unter: https://www.bamf.de/SharedDocs/Anlagen/DE/EMN/Politikberichte/emn-politikbericht-2016-germany.pdf?__blob=publicationFile&v=14 [22.02.2023].

Butt, Miriam / King, Tracy Holloway (2015): Lexical-Functional Grammar. In: Kiss, Tibor / Alexiadou, Artemis (Hg.): Syntax – Theory and Analysis. An International Handbook. Volume 2. Berlin / München / Boston: De Gruyter, 839–874. DOI: 10.1515/9783110363708–002.

Büning, Herbert / Trenkler, Götz (1994): Nichtparametrische statistische Methoden. Berlin / Boston: De Gruyter. DOI: 10.1515/9783110902990.

Byrnes, Heidi (2013): Notes from the editor. In: Modern Language Journal, 97/4, 825–827. DOI: 10.1111/j.1540–4781.2013.12051.x.

Byrnes, Heidi (2020): Envisioning L2 writing development in CDST under a curricular optic. In: Fogal, Gary / Verspoor, Marjolijn (Hg.): Complex Dynamic Systems Theory and L2 Writing Development. Amsterdam: John Benjamins, 241–270. DOI: 10.1075/lllt.54.10byr.

Cancino, Herlinda / Rosansky, Ellen J. / Schumann, John H. (1975): The Acquisition of the English Auxiliary by Native Spanish Speakers. In: TESOL Quarterly, 9/4, 421. DOI: 10.2307/3585626.

Carlo, Sabina de / Gamper, Jana (2015): Die Ermittlung grammatischer Kompetenzen anhand der Profilanalyse. In: Ziegler, Arne / Köpcke, Klaus-Michael (Hg.): Deutsche Grammatik in Kontakt. Berlin / Boston: De Gruyter, 103–136. DOI: 10.1515/9783110367171–006.

Carrion, Victor G. / Wong, Shane S. (2012): Can Traumatic Stress Alter the Brain? Understanding the Implications of Early Trauma on Brain Development and Learning. In: Journal of Adolescent Health, 51/2, 23–28. DOI: 10.1016/j.jadohealth.2012.04.010.

Caspari, Daniela / Klippel, Friederike / Legutke, Michael / Schramm, Karen (Hg.) (2016): Forschungsmethoden in der Fremdsprachendidaktik. Ein Handbuch. Tübingen: Narr Francke Attempto.

Chan, Hui-Ping (2015): A dynamic approach to the development of lexicon and syntax in a second language. Groningen: University of Groningen. Unter: https://research.rug.nl/en/publications/a-dynamic-approach-to-the-development-of-lexicon-and-syntax-in-a- [22.02.2023].

Chan, Hui-Ping / Verspoor, Marjolijn / Vahtrick, Louisa (2015): Dynamic Development in Speaking Versus Writing in Identical Twins. In: Language Learning, 65/2, 298–325. DOI: 10.1111/lang.12107.

Chen, Xiaobin B. / Meurers, Detmar (2016): CTAP: A Web-Based Tool Supporting Automatic Complexity Analysis. In: Proceedings of The Workshop on Computational Linguistics for Linguistic Complexity. Osaka, Japan. The International Committee on Computational Linguistics. Unter: https://

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

nebula.wsimg.com/b8835a55fa250ae67a999b1d57541c13?AccessKeyId=06FC08AB88B93C0DC287&disposition=0&alloworigin=1 [22.02.2023].
Chomsky, Noam (1965): Aspects of the theory of syntax. Cambridge: M.I.T. Press.
Cindark, Ibrahim / Deppermann, Arnulf / Hünlich, David / Lang, Christian / Perlmann-Balme, Michaela / Schöningh, Ingo (2019): Perspektive Beruf. Mündliche Kompetenz von Teilnehmenden an Integrationskursen und Vorschläge für die Praxis. Leibniz-Institut für Deutsche Sprache: Mannheim. Unter: http://dib.ids-mannheim.de/fileadmin/dateien/pdfs/Cindark_et_al__2019__Perspektive_Beruf.pdf [22.02.2023].
Cindark, Ibrahim / Hünlich, David (2019): Integrationskurse auf dem Prüfstand. Studie zu mündlichen Kompetenzen der Teilnehmenden an Integrationskursen. In: Sprachreport, 35/4, 18–25.
Clahsen, Harald (1982): Spracherwerb in der Kindheit. Eine Untersuchung zur Entwicklung der Syntax bei Kleinkindern. Tübingen: Narr.
Clahsen, Harald (1985): Profiling second language development: A procedure for assessing L2 proficiency. In: Hyltenstam, Keneth / Pienemann, Manfred (Hg.): Modelling and Assessing Second Language Acquisition. Clevedon: Multilingual Matters, 283–331.
Clahsen, Harald (1986): Die Profilanalyse. Ein linguistisches Verfahren für die Sprachdiagnose im Vorschulalter. Berlin: Marhold.
Clahsen, Harald / Meisel, Jürgen / Pienemann, Manfred (1983): Deutsch als Zweitsprache. Der Spracherwerb ausländischer Arbeiter. Tübingen: Narr.
Clyne, Michael (1968): Zum Pidgin-Deutsch der Gastarbeiter. In: Zeitschrift für Mundartenforschung, 35, 130–139.
Coenen, Anita (2017): Flucht – Trauma – Zweitsprache. Eine explorative Interviewstudie zu Trauma(-folgestörungen) bei Flüchtlingen und im Deutsch-als-Zweitsrpache-Unterricht. Masterarbeit. Jena: Friedrich-Schiller-Universität Jena. Unter: file:///C:/Users/ms/Downloads/bm_band_03_coenen_20170711.pdf.pdf [22.02.2023].
Collins, Linda (2006): Analysis of longitudinal data. The integration of theoretical model, temporal design, and statistical model. In: Annual review of psychology 57, 505–528. DOI: 10.1146/annurev.psych.57.102904.190146.
Corder, Pit (1973): Introducing Applied Linguistics. Harmondsworth: Penguin Education.
Crowther, Dustin / Kim, Susie / Lee, Jongbong / Lim, Jungmin / Loewen, Shawn (2020): Methodological Synthesis of Cluster Analysis in Second Language Research. In: Language Learning, 71/1, 99–130. DOI: 10.1111/lang.12428.
Csizér, Kata / Dörnyei, Zoltán (2005): Language learners' motivational profiles and their motivated learning behavior. In: Language Learning, 55/4, 613–659. DOI: 10.1111/j.0023–8333.2005.00319.x.
Cumming, Geoff (2014): The new statistics: why and how. In: Psychological Science, 25/1, 7–29. DOI: 10.1177/0956797613504966.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Czinglar, Christine (2014): Grammatikerwerb vor und nach der Pubertät. Berlin / Boston: De Gruyter. DOI: 10.1515/9783110332605.

Czinglar, Christine / Korecky-Kröll, Katharina / Uzunkaya-Sharma, Kumru / Dressler, Wolfgang U. (2015): Wie beeinflusst der sozioökonomische Status den Erwerb der Erst- und Zweitsprache? Wortschatzerwerb und Geschwindigkeit im NP/DP-Erwerb bei Kindergartenkindern im türkisch-deutschen Kontrast. In: Köpcke, Klaus-Michael (Hg.): Deutsche Grammatik in Kontakt. Deutsch als Zweitsprache in Schule und Unterricht. Berlin: De Gruyter, 207–240. DOI: 10.1515/9783110367171–010.

DeKeyser, Robert (2015): Skill acquisition theory. In: VanPatten, Bill / Williams, Jessica (Hg.): Theories in second language acquisition: An introduction. 2. Auflage. New York: Routledge, 94–112.

Demmig, Silvia (2018): Schreiben in der Zweitsprache Deutsch in der Erwachsenenbildung. In: Grießhaber, Wilhelm / Schmölzer-Eibinger, Sabine / Roll, Heike / Schramm, Karen (Hg.): Schreiben in der Zweitsprache Deutsch. Berlin / Boston: De Gruyter, 411–425. DOI: 10.1515/9783110354577–027.

Deppermann, Arnulf / Cindark, Ibrahim / Hünlich, David / Eichinger, Ludwig (Hg.) (2018): Flüchtlinge in Deutschland: Sprachliche und kommunikative Aspekte. Themenheft in Deutsche Sprache 3/2018. Berlin: Erich Schmidt.

Diehl, Erika (1998): Vom Deutschunterricht zum Deutscherwerb? Das Genfer DiGS-Projekt. In: Deutsch als Fremdsprache, 35/3, 162–168. DOI: 10.37307/j.2198–2430.1998.03.08.

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

Diehl, Erika / Pistorius, Hannelore / Fayolle Dietl, Annie (2002): Grammatikerwerb im Fremdsprachenunterricht – ein Widerspruch in sich? In: Börner, Wolfgang / Vogel, Klaus (Hg.): Grammatik und Fremdsprachenerwerb. Kognitive, psycholinguistische und erwerbstheoretische Perspektiven. Tübingen: Narr Francke Attempto, 143–163.

Diehl, Erika / Studer, Thérèse / Pistorius, Hannelore (2020): Grammatik im Gemeinsamen Europäischen Referenzrahmen? Erwerbssequenzen als Orientierungshilfe. In: Deutsch als Fremdsprache, 57/3, 143–154. DOI: 10.37307/j.2198–2430.2020.03.03.

Dirim, İnci / Krehut, Anne-Kristin (2020): Sprachgebrauch außerhalb der Schule. In: Ahrenholz, Bernt / Oomen-Welke, Ingelore / Ulrich, Winfried (Hg.): Deutsch als Zweitsprache. 5. unveränderte Auflage. Baltmannsweiler: Schneider Hohengehren, 535–547.

Dittmar, Norbert / Reich, Astrid / Schumacher, Magdalene / Skiba, Romuald / Terborg, Heiner (1990): Die Erlernung modaler Konzepte des Deutschen durch erwachsene polnische Migranten. Eine empirische Längsschnittstudie. In: Informationen Deutsch als Fremdsprache, 17/2, 125–172. DOI: 10.1515/infodaf–1990–170202.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Dolnicar, Sara (2002): Review of Unquestioned Standards in Using ClusterAnalysis for Data-driven Market Segmentation. CD Conference Proceedings of the Australian and NewZealand Marketing Academy Conference 2002 (ANZMAC 2002), Deakin University, Melbourne, 2.–4. Dezember 2002. Unter: http://ro.uow.edu.au/commpapers/273 [22.02.2023].

Domenech, Madeleine / Petersen, Inger (2018): Schriftliches Argumentieren. In: Grießhaber, Wilhelm / Schmölzer-Eibinger, Sabine / Roll, Heike / Schramm, Karen (Hg.): Schreiben in der Zweitsprache Deutsch. Berlin / Boston: De Gruyter, 108–120. DOI: 10.1515/9783110354577–008.

Döll, Marion (2009): Beobachtung und Dokumentation von Kompetenz und Kompetenzzuwachs im Deutschen als Zweitsprache mit den Niveaubeschreibungen DaZ. In: Lengyel, Dorit / Reich, Hans H. / Roth, Hans-Joachim / Döll, Marion (Hg.): Von der Sprachdiagnose zur Sprachförderung. Münster / New York: Waxmann, 109–114.

Döll, Marion (2009a): Das Instrument ‚Niveaubeschreibungen Deutsch als Zweitsprache für die Sekundarstufe I'. In: Sächsisches Bildungsinstitut (Hg.): Niveaubeschreibungen Deutsch als Zweitsprache für die Sekundarstufe I. Radebeul: Sächsisches Bildungsinstitut, 11–15.

Dörnyei, Zoltán / MacIntyre, Peter D. / Henry, Alastair (Hg.) (2015): Motivational dynamics in language learning. Bristol: Multilingual Matters. DOI: 10.21832/9781783092574.

Draganski, Bogdan / Gaser, Christian / Busch, Volker / Schuierer, Gerhard / Bogdahn, Ulrich / May, Arne (2004): Neurplasticity: Changes in grey matter induced by training. In: Nature, 427, 311–312. DOI: 10.1038/427311a.

Drumm, Sandra / Henning, Ute (2016): Bildungs- und Fachsprache als Herausforderung und Ziel der sprachlichen Bildung von Fluchtmigrantinnen und -migranten. In: Fremdsprache Deutsch. Zeitschrift für die Praxis des Deutschunterrichts. Sonderheft: Deutschunterricht für Lernende mit Migrationshintergrund, 11–14.

Duff, Patricia A. (2014): Case Study Research on Language Learning and Use. In: Annual Review of Applied Linguistics, 34, 233–255. DOI: 10.1017/S0267190514000051.

Dulay, Heidi / Burt, Marina (1973): Should we teach Children Syntax? In: Language Learning, 23/2, 245–258. DOI: 10.1111/j.1467–1770.1973.tb00659.x.

Dulay, Heidi / Burt, Marina (1974): Natural sequences in child second language acquisition. In: Language Learning, 24/1, 37–53. DOI: 10.1111/j.1467–1770.1974.tb00234.x.

Dürscheid, Christa (2012): Syntax. Grundlagen und Theorien. Göttingen: Vandenhoek & Ruprecht.

Dyson, Bronwen Patricia (2021): Dynamic Variation in Second Language Acquisition. Amsterdam: John Benjamins. DOI: 10.1075/palart.8.

Dyson, Bronwen (2022): SLA as complex, dynamical and predictable: A Processability Theory perspective. In: Second Language Research, Online First. DOI: 10.1177/02676583221132726.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Efron, Bradley / Tibshirani, Robert J. (1994): An Introduction to the Bootstrap. New York: Chapman and Hall/CRC. DOI: 10.1201/9780429246593.

Ehl, Birgit / Paul, Michèle / Bruns, Gunnar / Fleischhauer, Elisabeth / Vock, Miriam / Gronostaj, Anna / Grosche, Michael (2018): Testgütekriterien der „Profilanalyse nach Grießhaber". Evaluation eines Verfahrens zur Erfassung grammatischer Fähigkeiten von ein- und mehrsprachigen Grundschulkindern. In: Zeitschrift für Erziehungswissenschaften, 21/6, 1261–1281. DOI: 10.1007/s11618–018–0835-x.

Ehlich, Konrad (Hg.) (2007): Sprache und sprachliches Handeln. Band 1. Berlin / New York: De Gruyter. DOI: 10.1515/9783110922721.

Ehlich, Konrad (2007a): Funktional-pragmatische Kommunikationsanalyse: Ziele und Verfahren. In: Ehlich, Konrad (Hg.): Sprache und sprachliches Handeln. Band 1. Berlin / New York: De Gruyter, 9–28. DOI: 10.1515/9783110922721.1.9.

Eichinger, Ludwig M. (2018): Korpuslinguistik im Alltag linguistischer Forschung. In: Osnabrücker Beiträge zur Sprachtheorie, 92, 69–82.

Ellis, Nick (2005): At the interface: Dynamic interactions of explicit and implicit language knowledge. In: Studies in Second Language Acquisition, 27/2, 305–352. DOI: 10.1017/S027226310505014X.

Ellis, Nick (2007): The weak interface, consciousness, and form-focused instruction: mind the doors. In: Fotos, Sandra / Hossein, Nassaj (Hg.): Form Focused instruction and Teacher Education. Oxford: Oxford University Press.

Ellis, Rod (1989): Are Classroom and Naturalistic Acquisition the Same? In: Studies in Second Language Acquisition, 11/3, 305–328. DOI: 10.1017/S0272263100008159.

Europarat (Hg.) (2001): Gemeinsamer europäischer Referenzrahmen für Sprachen. Lernen, lehren, beurteilen [Niveau A1, A2, B1, B2, C1, C2]. München: Klett-Langenscheidt.

Evans, Sandra / Pude, Angela / Specht, Franz (2012): Menschen A1. Kursbuch. Ismaning: Hueber.

Fandrych, Christian (2003): Zur Textlinguistik des Vorfelds. In: Thurmair, Maria / Willkop, Eva-Maria (Hg.): Am Anfang war der Text. 10 Jahre „Textgrammatik der deutschen Sprache". München: iudicium, 173–196.

Fandrych, Christian (Hg.) (2005): Ordnung und Variation in Satz und Text. Themenheft Fremdsprache Deutsch 32. Stuttgart: Klett.

Fandrych, Christian (2013): Grammatik. In: Ahrenholz, Bernt / Oomen-Welke, Ingelore (Hg.): Deutsch als Fremdsprache. Baltmansweiler: Schneider Hohengehren, 165–177.

Fandrych, Christian / Tschirner, Erwin (2007): Korpuslinguistik und Deutsch als Fremdsprache. In: Deutsch als Fremdsprache, 44/4, 195–204. DOI: 10.37307/j.2198–2430.2007.04.02.

Fandrych, Christian / Thurmair, Maria (2011): Textsorten im Deutschen. Linguistische Analysen aus sprachdidaktischer Sicht. Tübingen: Stauffenburg.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Fandrych, Christian / Thurmair, Maria (2016): Grammatik an und mit Textsorten lernen: Das Prinzip der „Linksdeterminierung" im Deutschen. In: Freudenberg-Findeisen, Renate (Hg.): Auf dem Weg zu einer Textsortendidaktik. Linguistische Analysen und text(sorten)didaktische Bausteine nicht nur für den fremdsprachlichen Deutschunterricht. Hildesheim: Olms, 185–200.

Fandrych, Christian / Wallner, Franziska (2022): Funktionale und stilistische Merkmale gesprochener fortgeschrittener Lerner:innensprache: Methodische und konzeptionelle Überlegungen am Beispiel von GeWiss. In: Zeitschrift für germanistische Linguistik, 50/1, 202–239.

Fandrych, Christian / Hufeisen, Britta / Mohr, Imke / Thonhauser, Ingo / Wicke, Rainer E. / Dronske, Ulrich (Hg.) (2016): Deutschunterricht für Lernende mit Migrationshintergrund. Berlin: Erich Schmidt (Fremdsprache Deutsch Sonderheft, 2016).

Feick, Diana (2018): Diagnostik in der Erwachsenenbildung. In: Grießhaber, Wilhelm / Schmölzer-Eibinger, Sabine / Roll, Heike / Schramm, Karen (Hg.): Schreiben in der Zweitsprache Deutsch. Berlin / Boston: De Gruyter, 247–264. DOI: 10.1515/9783110354577–017.

Feilke, Helmut (2015): Text und Lernen – Perspektivenwechsel in der Schreibforschung. In: Schmölzer-Eibinger, Sabine / Thürmann, Eike (Hg.): Schreiben als Medium des Lernens. Kompetenzentwicklung durch Schreiben. Münster: Waxmann, 47–71. DOI: 10.1515/9783110354577–017.

Field, Andy (2018): Discovering statistics using IBM SPSS statistics. Los Angeles / London / New Delhi, Singapore / Washington DC / Melbourne: SAGE.

Fogal, Gary / Verspoor, Marjolijn (Hg.) (2020): Complex Dynamic Systems Theory and L2 Writing Development. Amsterdam: John Benjamins. DOI: 10.1075/lllt.54.

Fogal, Gary / Verspoor, Marjolijn (2020a): Introduction. In: Fogal, Gary / Verspoor, Marjolijn (Hg.): Complex Dynamic Systems Theory and L2 Writing Development. Amsterdam: John Benjamins, ix–xvii. DOI: 10.1075/lllt.54.int.

Formann, Anton (1984): Die Latent-Class-Analyse: Einführung in die Theorie und Anwendung. Weinheim: Beltz.

Fotos, Sandra / Hossein, Nassaj (Hg. 2007): Form Focused instruction and Teacher Education. Oxford: Oxford University Press.

Friederici, Angela D. (2011): The Brain Basis of Language Processing: From Structure to Function. In: Physiological Reviews, 91/4, 1357–1392. DOI: 10.1152/physrev.00006.2011.

Friederici, Angela D. (2017): Language in our brain. The origins of a uniquely human capacity. Mit einem Vorwort von Noam Chomsky. Cambridge / London: The MIT Press. DOI: 10.7551/mitpress/11173.001.0001.

Friederici, Angela D. / Gierhan, Sarah (2013): The language network. In: Current Opinion in Neurobiology, 23/2, 250–254. DOI: 10.1016/j.conb.2012.10.002.

Friedman, Milton (1937): The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. In: Journal of the American Statistical Association, 32/200, 675–701. DOI: 10.1080/01621459.1937.10503522.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Freudenberg-Findeisen, Renate (Hg.) (2016): Auf dem Weg zu einer Textsortendidaktik. Linguistische Analysen und text(sorten)didaktische Bausteine nicht nur für den fremdsprachlichen Deutschunterricht. Hildesheim: Olms.

Fürstenau, Sara / Gomolla, Mechtild (Hg.) (2012): Migration und schulischer Wandel: Leistungsbeurteilung. Wiesbaden: Springer.

Gaillard, Stéphanie / Tremblay, Annie (2016): Linguistic Proficiency Assessment in Second Language Acquisition Research: The Elicited Imitation Task. In: Language Learning, 66/2, 419–447. DOI: 10.1111/lang.12157.

Gallmann, Peter (2016): Wortstellung: die Abfolge von Satzgliedern und Prädikatsteilen im Satz. In: Wöllstein, Angelika / Dudenredaktion (Hg.): Duden. Die Grammatik. Unentbehrlich für richtiges Deutsch. 9., vollständige überarbeitete und aktualisierte Auflage. Berlin: Bibliographisches Institut, 871–904.

Gamper, Jana / Schroeder, Christoph (2016): Sprachliche Bildung für Neuzugewanderte. Ein Plädoyer für einen erwerbssequentiellen Ansatz. In: OBST (Flucht_Punkt_Sprache), 89, 217–230.

Gamper, Jana / Marx, Nicole / Röttger, Evelyn / Steinbock, Dorotheé (2020): Deutsch für Seiteneinsteiger/innen. In: Informationen Deutsch als Fremdsprache, 47/4, 347–358. DOI: 10.1515/infodaf–2020–0067.

Gamper, Jana / Schroeder, Christoph (2021): Lehr- und Lernkontexte des Deutschen als Zweitsprache in den deutschsprachigen Ländern. In: Altmayer, Claus / Biebighäuser, Katrin / Haberzettl, Stefanie / Heine, Antje (Hg.) (2021): Handbuch Deutsch als Fremd- und Zweitsprache. Stuttgart: J.B. Metzler, 63–76. DOI: 10.1007/978–3–476–04858–5_5.

Geist, Barbara (2021): Migrationsbedingte Mehrsprachigkeit. In: Altmayer, Claus / Biebighäuser, Katrin / Haberzettl, Stefanie / Heine, Antje (Hg.): Handbuch Deutsch als Fremd- und Zweitsprache. Stuttgart: J.B. Metzler, 77–85. DOI: 10.1007/978–3–476–04858–5_6.

Geertzen, Jeroen / Alexopoulou, Theodora / Korhonen, Anna (2013): Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge Open Language Database (EFCAMDAT). Selected Proceedings of the 31st Second Language Research Forum (SLRF). Cascadilla.

Glaboniat, Manuela / Perlmann-Balme, Michaela / Studer, Thomas (2013): Zertifikat B1. Deutschprüfung für Jugendliche und Erwachsene: Prüfungsziele, Testbeschreibung. Ismaning: Hueber.

Glas-Peters, Sabine / Pude, Angela / Reimann, Monika (2012): Menschen A1. Arbeitsbuch. Ismaning: Hueber.

Gogolin, Ingrid (2004): Lebensweltliche Mehrsprachigkeit. In: Bausch, Karl-Richard (Hg.): Mehrsprachigkeit im Fokus. Arbeitspapiere der 24. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts. Tübingen: Narr, 55–61.

Gogolin, Ingrid / Dirim, İnci / Klinger, Thorsten / Lange, Imke / Lengyel, Drorit / Neumann, Ursula / Michel, Ute / Reich, Hans H. / Schwippert, Knut (2011): Förderung von Kindern und Jugendlichen mit Migrationshintergrund FörMig. Bilanz und Perspektiven eines Modellprogramms. Münster: Waxmann.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Goschler, Juliana (2017): Syntaxerwerb bei ein- und mehrsprachigen Grundschüler/innen: eine quantitative Untersuchung. In: Zeitschrift für angewandte Linguistik, 66/1. DOI: 10.1515/zfal–2017–0002.

Gregg, Kevin R. (2010): Review article: Shallow draughts: Larsen-Freeman and Cameron on complexity. In: Second Language Research, 26/4, 549–560. DOI: 10.1177/0267658310366582.

Gries, Stefan Th. (2021): (Generalized Linear) Mixed-Effects Modeling: A Learner Corpus Example. In: Language Learning, 71/3, 757–798. DOI: 10.1111/lang.12448.

Grießhaber, Wilhelm (2001): Erst- und zweitsprachliche Textproduktion nach Bildvorlage. In: Wolff, Armin / Winters-Ohle, Elmar (Hg.): Wie schwer ist die deutsche Sprache wirklich? Regensburg: Fachverband Deutsch als Fremdsprache, 102–114.

Grießhaber, Wilhelm (2001a): Erwerb und Vermittlung des Deutschen als Zweitsprache. In: „Deutsch in Armenien" Teil 1: 1/2001,17–24; Teil 2 2/2001. Jerewan: Armenischer Deutschlehrerverband, 5–15.

Grießhaber, Wilhelm (2005): Sprachstandsdiagnose im kindlichen Zweitspracherwerb: Funktional-pragmatische Fundierung der Profilanalyse. Münster: WWU Sprachenzentrum. Veröffentlicht auf dem IDS-Publikationsserver. DOI: 10.14618/ids-pub–10388.

Grießhaber, Wilhelm (2006): Lernende unterstützen: die Profilanalyse als didaktisch nutzbares Werkzeug der Lernersprachenanalyse. Münster: WWU Sprachenzentrum. Unter: https://www.yumpu.com/de/document/read/9160112/wilhelm-griesshaber-lernende-unterstutzen-die-profilanalyse-als- [22.02.2023].

Grießhaber, Wilhelm (2007): „und wir faren in die andere seite" – Der Gebrauch lokaler Präpositionen durch türkische Grundschüler. In: Meng, Katharina / Rehbein, Jochen (Hg.): Kindliche Kommunikation – einsprachig und mehrsprachig. Münster: Waxmann, 371–392.

Grießhaber, Wilhelm (2012): Die Profilanalyse. In: Ahrenholz, Bernt (Hg.): Einblicke in die Zweitspracherwerbsforschung und ihre methodischen Verfahren. Berlin / Boston: De Gruyter. DOI: 10.1515/9783110267822.173.

Grießhaber, Wilhelm (2014): Beurteilung von Texten mehrsprachiger Schülerinnen und Schüler. In: leseforum.ch, 3, 1–19. Unter: http://www.leseforum.ch/myUploadData/files/2014_3_Griesshaber.pdf [22.02.2023].

Grießhaber, Wilhelm (2016): Spracherwerbsprozesse in Erst- & Zweitsprache. Eine Einführung. 3., unveränderte Auflage. Duisburg: Universitätsverlag Rhein-Ruhr.

Grießhaber, Wilhelm (2018): Zweitspracherwerb und Diagnose des Schreibens in der Zweitsprache. In: Grießhaber, Wilhelm / Schmölzer-Eibinger, Sabine / Roll, Heike / Schramm, Karen (Hg.): Schreiben in der Zweitsprache Deutsch. Berlin / Boston: De Gruyter, 17–32. DOI: 10.1515/9783110354577–002.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Grießhaber, Wilhelm (2020): Schreiben in der Zweitsprache Deutsch. In: Ahrenholz, Bernt / Oomen-Welke, Ingelore / Ulrich, Winfried (Hg.): Deutsch als Zweitsprache. 5. unveränderte Auflage. Baltmannsweiler: Schneider Hohengehren, 292–305.

Grießhaber, Wilhelm / Schmölzer-Eibinger, Sabine / Roll, Heike / Schramm, Karen (Hg.) (2018): Schreiben in der Zweitsprache Deutsch. Berlin / Boston: De Gruyter. DOI: 10.1515/9783110354577.

Grimm, Angela / Schulz, Petra (2012): Forschungsmethoden der kombinierten Längs- und Querschnittsstudie MILA. In: Ahrenholz, Bernt (Hg.): Einblicke in die Zweitspracherwerbsforschung und ihre methodischen Verfahren. Berlin / Boston: De Gruyter, 195–218. DOI: 10.1515/9783110267822.195.

Gudmestad, Aarnes / Edmonds, Amanda (2018): Critical Reflections on Data in Second Language Acquisition. Amsterdam: John Benjamins. DOI: 10.1075/lllt.51.

Guessoum, Nidhal (2012): Does the Arab world (not) need basic science? Nature Middle East. DOI: doi:10.1038/nmiddleeast.2012.52.

Goucha, Tomás / Anwander, Alfred / Friederici, Angela D. (2015): How language shapes the brain: Cross-linguistic differences in structural connectivity. 7th Annual Meeting of the Society for the Neurobiology of Language, Chicago, IL, USA. [Poster].

Guijarro-Fuentes, Pedro / Suarez-Gomez, Cristina / Vulchanova, Mila / Sorace, Antonella / Vulchanov, Valentin (Hg.) (2021|2022): Research Topic: The Notion of the Native Speaker Put to the Test: Recent Research Advances. In: Frontiers in Psychology. Unter: https://www.frontiersin.org/research-topics/13914/the-notion-of-the-native-speaker-put-to-the-test-recent-research-advances#overview [22.2.2023].

Gültekin-Karakoç, Nazan / Feldmeier, Alexis (2014): Analyse Quantitativer Daten. In: Settinieri, Julia / Demirkaya, Sevilen / Feldmeier, Alexis / Gültekin-Karakoç, Nazan / Riemer, Claudia (Hg.): Empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache. Eine Einführung. Paderborn: Ferdinand Schöningh, 183–211.

Gürsoy, Erkan (2010): Komplexe Nominalgruppen im Deutschen und in weiteren Sprachen. In: Kompetenzzentrum ProDaZ. Unter: http://www.uni-due.de/imperia/md/content/prodaz/komplexe_nominalgruppen.pdf [2.10.2020].

Habersack, Charlotte / Pude, Angela / Specht, Franz (2019): Menschen A2. Kursbuch. München: Hueber.

Haberzettl, Stefanie (2005): Der Erwerb der Verbstellungsregeln in der Zweitsprache Deutsch durch Kinder mit russischer und türkischer Muttersprache. Berlin / Boston: De Gruyter. DOI: 10.1515/9783110922127.

Håkansson, Gisela (2001): Against Full Transfer evidence from Swedish learners of German. Unter: https://lucris.lub.lu.se/ws/files/6074585/624434.pdf [22.02.2023].

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Håkansson, Gisela / Pienemann, Manfred / Sayehli, Susan (2002): Transfer and Typological Proximity in the Context of Second Language Processing. In: Second Language Research, 18/3, 250–73.

Han, ZaoHong (2004): To Be a Native Speaker Means Not to Be a Nonnative Speaker. In: Second Language Research, 20/2, 166–87. DOI: 10.1191/0267658304sr235oa.

Hawkins, John A. / Filipović, Luna (2012): Criterial features in L2 English. Specifying the reference levels of the Common European Framework. Cambridge: Cambridge University Press.

Hayes, John R. / Flower, Linda S. (1980): Identifying the organization of writing processes. In: Gregg, Lee / Steinberg, Erwin Ray (Hg): Cognitive processes in writing. Hillsdale: Erlbaum, 3–30.

Han, ZhaoHong / Tarone, Elaine (Hg.) (2014): Interlanguage: 40 Years later. Amsterdam: John Benjamins. DOI: 10.1075/lllt.39.

Heidelberger Forschungsprojekt „Pidgin-Deutsch" (HDP) (1977): Heidelberger Forschungsprojekt „Pidgin-Deutsch spanischer und italienischer Arbeiter in der Bundesrepublik": Die ungesteuerte Erlernung des Deutschen durch spanische und italienische Arbeiter. Eine soziologische Untersuchung. In: Osnabrücker Beiträge zur Sprachtheorie, Beiheft 2.

Heilmann, Beatrix (2012): Diagnostik & Förderung – leicht gemacht. Stuttgart: Ernst Klett Sprachen.

Hernandez, Arturo (2013): The bilingual brain. Oxford: Oxford University Press. DOI:10.1093/acprof:oso/97801998281111.001.0001.

Hinger, Barbara / Stadler, Wolfgang (2018): Testen und Bewerten fremdsprachlicher Kompetenzen im Unterricht. Eine Einführung. Tübingen: Narr Francke Attempto.

Hirschmann, Hagen (2019): Korpuslinguistik. Eine Einführung. Stuttgart: J.B. Metzler. DOI: 10.1007/978–3–476–05493–7.

Hiver, Phil / Al-Hoorie, Ali H. (2019): Research Methods for Complexity Theory in Applied Linguistics: Multilingual Matters. DOI: 10.21832/HIVER5747.

Hiver, Phil / Al-Hoorie, Ali H. / Evans, Reid (2021): Complex Dynamic Systems Theory in Language Learning. In: Studies in Second Language Acquisition, 1–29. DOI: 10.1017/S0272263121000553.

Hou, Junping / Loerts, Hanneke / Verspoor, Marjolijn (2020): Coordination of linguistic subsystems as a sign of automatization? In: Fogal, Gary / Verspoor, Marjolijn (Hg.): Complex Dynamic Systems Theory and L2 Writing Development. Amsterdam: John Benjamins, 27–48. DOI: 10.1075/lllt.54.02hou.

Housen, Alex, / Kuiken, Folkert / Vedder, Ineke (Hg.) (2012): Dimensions of L2 Performance and Proficiency. Amsterdam: John Benjamins. DOI: 10.1075/lllt.32.

Housen, Alex / Kuiken, Folkert / Vedder, Ineke (2012a): Complexity, accuracy and fluency: Definitions, measurement and research. In: Housen, Alex, / Kuiken, Folkert / Vedder, Ineke (Hg.): Dimensions of L2 Performance and Proficiency. Amsterdam: John Benjamins, 1–20. DOI: 10.1075/lllt.32.01hou.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Housen, Alex / Clercq, Bastien de / Kuiken, Folkert / Vedder, Ineke (2019): Multiple approaches to complexity in second language research. In: Second Language Research, 35/1, 3–21. DOI: 10.1177/0267658318809765.

Huang, Yan / Murakami, Akira / Alexopoulou, Theodora / Korhonen, Anna (2018): Dependency parsing of learner English. In: International Journal of Corpus Linguistics, 23/1, 28–54. DOI: 10.1075/ijcl.16080.hua.

Hufeisen, Britta (2003): L1, L2, L3, L4, Lx – alle gleich? Linguistische, lernerinterne und lernerexterne Faktoren in Modellen zum multiplen Spracherwerb. In: Zeitschrift für Interkulturellen Fremdsprachenunterricht, 8/2–3, 97–109. Unter: https://ojs.tujournals.ulb.tu-darmstadt.de/index.php/zif/article/view/537/513 [22.2.2023].

Hufeisen, Britta / Riemer, Claudia (2010): Spracherwerb und Sprachenlernen. In: Krumm, Hans-Jürgen / Fandrych, Christian / Hufeisen, Britta / Riemer, Claudia (Hg.): Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch. 1. Halbband. Berlin / New York: De Gruyter, 738–753. DOI: 10.1515/9783110240245.738.

Hulstijn, Jan H. / Ellis, Ro / Eskildsen, Søren W. (2015): Orders and Sequences in the Acquisition of L2 Morphosyntax, 40 Years On: An Introduction to the Special Issue. In: Language Learning, 65/1, 1–5. DOI: 10.1111/lang.12097.

Hünlich, David / Wolfer, Sascha / Lang, Christian / Deppermann, Arnulf (2018): Wer besucht den Integrationskurs? Soziale und sprachliche Hintergründe von Geflüchteten und anderen Zugewanderten. Mannheim: Institut für Deutsche Sprache; Goethe-Institut. Unter: https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/7668/file/Huenlich_etal_Wer_besucht_den_Integrationskurs_2018.pdf [22.02.2023].

Hyltenstam, Keneth / Pienemann, Manfred (Hg.) (1985): Modelling and Assessing Second Language Acquisition. Clevedon: Multilingual Matters.

Institut für Testforschung und Testentwicklung (ITT) (2017): Test grammatischer Kompetenz Deutsch als Zweitsprache (GraKO-DaZ) Beschreibung. Bis jetzt nicht veröffentlicht.

Ioannidis, John P. A. (2005): Why most published research findings are false. In: PLoS Medicine, 2/8, e124, 696–701. DOI: 10.1371/journal.pmed.0020124.

Isbell, Dan / Winke, Paula (2019): ACTFL Oral Proficiency Interview – computer (OPIc). In: Language Testing, 49/4, DOI: 10.1177/0265532219828253, 1–11.

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. DOI: 10.1111/j.1467–9922.2007.00438.x.

JASP Team (2020): JASP (Version 0.14.1) [Computer software].

Jeuk, Stefan (2018): Schriftspracherwerb und Alphabetisierung in der Zweitsprache im Grundschulalter. In: Grießhaber, Wilhelm / Schmölzer-Eibinger, Sabine / Roll, Heike / Schramm, Karen (Hg.): Schreiben in der Zweitsprache Deutsch. Berlin / Boston: De Gruyter, 49–62. DOI: 10.1515/9783110354577–004.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Johnsen, Grethe E. / Asbjørnsen, Arve E. (2009): Verbal learning and memory impairments in posttraumatic stress disorder: the role of encoding strategies. In: Psychiatry research, 165/1–2, 68–77. DOI: 10.1016/j.psychres.2008.01.001.

Kadamani, Miral (2012): Eine sprachpraktische Analyse zur Beurteilung der Schreibkompetenz arabischer Grundschulkinder im Deutschen: Inwiefern wirkt sich das Sprachwahlverhalten der Eltern auf die Schriftsprache der Kinder im Deutschen aus? Dissertation. Berlin. DOI: 10.14279/depositonce–3568.

Kaplan, Ina / Petersen, Inger (Hg.) (2019): Schreibkompetenzen messen, beurteilen und fördern. Münster / New York: Waxmann.

Keßler, Jörg-Uwe / Liebner, Mathias (2011): Diagnosing L2 development. In: Pienemann, Manfred / Keßler, Jörg-Uwe (Hg.): Studying Processability Theory. Amsterdam: John Benjamins, 131–147. DOI: 10.1075/palart.1.11dia.

Kiss, Tibor / Alexiadou, Artemis (Hg.) (2015): Syntax – Theory and Analysis. An International Handbook. Volume 2. Berlin / München / Boston: De Gruyter. DOI: 10.1515/9783110363708.

Klein Gunnewiek, Lisanne (2000): Sequenzen und Konsequenzen. Zur Entwicklung niederländischer Lerner im Deutschen als Fremdsprache. Amsterdam: Atlanta.

Klinger, Thorsten / Schwippert, Knut / Leiblein, Birgit (Hg.) (2008): Evaluation im Modellprogramm FörMig: Planung und Realisierung eines Evaluationskonzepts. Münster / New York.

Knappik, Magdalena / Dirim, İnci (2018): Diagnose zweitsprachlichen Schreibens. In: Grießhaber, Wilhelm / Schmölzer-Eibinger, Sabine / Roll, Heike / Schramm, Karen (Hg.): Schreiben in der Zweitsprache Deutsch. Berlin / Boston: De Gruyter, 167–182. DOI: 10.1515/9783110354577–012.

Knorr, Dagmar (2018): Schreibentwicklung in der Hochschule: wissenschaftliche Textkompetenz in der Zweitsprache Deutsch. In: Grießhaber, Wilhelm / Schmölzer-Eibinger, Sabine / Roll, Heike / Schramm, Karen (Hg.): Schreiben in der Zweitsprache Deutsch. Berlin / Boston: De Gruyter, 135–148. DOI: 10.1515/9783110354577–010.

Königs, Frank G. (2010): Zweitsprachenerwerb und Fremdsprachenlernen: Begriffe und Konzepte. In: Krumm, Hans-Jürgen / Fandrych, Christian / Hufeisen, Britta / Riemer, Claudia (Hg.): Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch. 1. Halbband. Berlin / New York: De Gruyter, 754–763. DOI: 10.1515/9783110240245.738.

Köpcke, Klaus-Michael (Hg.) (2015): Deutsche Grammatik in Kontakt. Deutsch als Zweitsprache in Schule und Unterricht. Berlin: De Gruyter. DOI: 10.1515/9783110367171.

Kramsch, Claire (2012): Why is everybody so excited about complexity theory in applied linguistics? In: Mélanges CRAPEL 33 (numero special: Didactique des langues et complexité: En hommage à Richard Duda [Sonderausgabe: Sprachunterricht und Komplexität: Eine Hommage an Richard Duda; [Übersetzung MS]), 9–24.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Krumm, Hans-Jürgen / Fandrych, Christian / Hufeisen, Britta / Riemer, Claudia (Hg.) (2010): Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch. 1. Halbband. Berlin / New York: De Gruyter. DOI: 10.1515/9783110240245.

Kuhs, Katharina (2020): Einflussfaktoren auf die schulische L2-Kompetenz von Schülerinnen und Schülern mit Deutsch als Zweitsprache. In: Ahrenholz, Bernt / Oomen-Welke, Ingelore / Ulrich, Winfried (Hg.): Deutsch als Zweitsprache. 5. unveränderte Auflage. Baltmannsweiler: Schneider Hohengehren, 518–531.

Larsen-Freeman, Diane (1997): Chaos/Complexity Science and Second Language Acquisition. In: Applied Linguistics, 18/2, 141–165. DOI: 10.1093/applin/18.2.141.

Larsen-Freeman, Diane (2006): The Emergence of Complexity, Fluency, and Accuracy in the Oral and Written Production of Five Chinese Learners of English. In: Applied Linguistics, 27/4, 590–619. DOI: 10.1093/applin/aml029.

Larsen-Freeman, Diane (2007): On the complementarity of Chaos/Complexity Theory and Dynamic Systems Theory in understanding the second language acquisition process. In: Bilingualism, 10/1, 35–37. DOI: 10.1017/S136672890600277X.

Larsen-Freeman, Diane (2014): Another step to be taken – Rethinking the end point of the interlanguage continuum. In: Han, ZhaoHong / Tarone, Elaine (Hg.): Interlanguage: 40 Years later. Amsterdam: John Benjamins, 203–220. DOI: 10.1075/lllt.39.11ch9.

Larsen-Freeman, Diane / Cameron, Lynne (2008): Research Methodology on Language Development from a Complex Systems Perspective. In: The Modern Language Journal, 92/2, 200–213. DOI: 10.1111/j.1540–4781.2008.00714.x.

Larson-Hall, Jenifer / Plonsky, Luke (2015): Reporting and Interpreting Quantitative Research Findings: What Gets Reported and Recommendations for the Field. In: Language Learning, 65/1, 127–159. DOI: 10.1111/lang.12115.

Landesamt für Schule und Bildung Sachsen [LASUB Sachsen] (2019): Niveaubeschreibungen Deutsch als Zweitsprache für die Sekundarstufe 1. Unter: https://publikationen.sachsen.de/bdb/artikel/14477 [22.02.2023].

Lauterbach, Stefan (2005): Wortstellung hören! – Ja, kann man das? Zum Aufbau grammatischer Hörmuster. In: Fandrych, Christian (Hg.): Ordnung und Variation in Satz und Text. Themenheft Fremdsprache Deutsch 32. Stuttgart: Klett.

Lengyel, Drorit / Reich, Hans H. / Roth, Hans-Joachim / Döll, Marion (Hg.) (2009): Von der Sprachdiagnose zur Sprachförderung. Münster / New York: Waxmann.

Lengyel, Drorit / Roth, Hans-Joachim (2012): Beobachtung der Schreibentwicklung in der Sekundarstufe I. In: Fürstenau, Sara / Gomolla, Mechtild (Hg.): Migration und schulischer Wandel: Leistungsbeurteilung. Wiesbaden: Springer, 123–136.

Lenzhofer-Glantschnig, Melanie / Scherr, Elisabeth (2015): Abweichungen sind keine Fehler. Reflexe innersprachlicher Variation im Unterricht für Deutsch

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

als Zweitsprache. In: Köpcke, Klaus-Michael (Hg.): Deutsche Grammatik in Kontakt. Deutsch als Zweitsprache in Schule und Unterricht. Berlin: De Gruyter, 153–177. DOI: 10.1515/9783110367171–008.

Lenzing, Anke (2015): Exploring Regularities and Dynamic Systems in L2 Development. In: Language Learning, 65/1, 89–122. DOI: 10.1111/lang.12092.

Lenzing, Anke / Nicholas, Howard / Roos, Jana (Hg.) (2019): Widening Contexts for Processability Theory. Amsterdam: John Benjamins. DOI: 10.1075/palart.7.

Lenzing, Anke / Nicholas, Howard / Roos, Jana (2019a): Contextualising issues in Processability Theory. In: Lenzing, Anke / Nicholas, Howard / Roos, Jana (Hg.): Widening Contexts for Processability Theory. Amsterdam: John Benjamins, 1–8. 10.1075/palart.7.01len.

Levelt, Willem J. M. (1989): Speaking. From intention to articulation. Cambridge: MIT.

Li, Yuan / Wu, Zekun (2022): Zur syntaktischen Komplexität des Schriftdeutschen chinesischer Deutschlerner/-innen – Eine korpusbasierte Profilanalyse. In: Deutsch als Fremdsprache, 59/3, 207–217. DOI: 10.37307/j.2198–2430.2022.04.04.

Linnemann, Markus (2017): Erfassung von Schreibprozessen: Methoden, Techniken, Tools. In: Becker-Mrotzek, Michael / Grabowski, Joachim / Steinhoff, Torsten (Hg.): Forschungshandbuch empirische Schreibdidaktik. Münster: Waxmann, 335–268.

Little, Roderick J. / Rubin, Donald B. (2002): Statistical analysis with missing data. Hoboken: Wiley. DOI:10.1002/9781119013563.

Lowie, Wander / Verspoor, Marjolijn (2015): Variability and Variation in Second Language Acquisition Orders: A Dynamic Reevaluation. In: Language Learning, 65/1, 63–88. DOI: 10.1111/lang.12093.

Lu, Xiaofei (2010): Automatic analysis of syntactic complexity in second language writing. In: International Journal of Corpus Linguistics, 15/4, 474–496. DOI: 10.1075/ijcl.15.4.02lu.

Lu, Xiaofei (2011): A Corpus-Based Evaluation of Syntactic Complexity Measures as Indices of College-Level ESL Writers' Language Development. In: TESOL Quarterly, 45/1, 36–62. DOI: 10.5054/tq.2011.240859.

Lund, Randall J. (2004): Erwerbssequenzen im Klassenraum. In: Deutsch als Fremdsprache, 41/2, 99–103. DOI: 10.37307/j.2198–2430.2004.02.06.

Lüdeling, Anke / Doolittle, Seanna / Hirschmann, Hagen / Schmidt, Karin / Walter, Maik (2008): Das Lernerkorpus Falko. In: Deutsch als Fremdsprache, 45/2, 67–73. DOI: 10.37307/j.2198–2430.2008.02.02.

Maas, Utz (2015): Sprachausbau in der Zweitsprache. In: Köpcke, Klaus-Michael (Hg.): Deutsche Grammatik in Kontakt. Deutsch als Zweitsprache in Schule und Unterricht. Berlin: De Gruyter, 1–23. DOI: 10.1515/9783110367171–002.

MacIntyre, Peter D. / MacKay, Emily / Ross, Jessica / Abel, Esther (2017): The emerging need for methods appropriate to study dynamic systems. In: Ortega, Lourdes / Han, ZhaoHong (Hg.): Complexity Theory and Language

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Development. In Celebration of Diane Larsen-Freeman. Amsterdam: John Benjamins, 97–122. DOI: 10.1075/lllt.48.06mac.

Macqueen, Susy (2012): The emergence of patterns in second language writing. A sociocognitive exploration of lexical trails. Bern: Peter Lang.

Marsden, Emma / Plonsky, Luke (2018): Data, open science, and methodological reform in second language acquisition. In: Gudmestad, Aarnes / Edmonds, Amanda (Hg.): Critical Reflections on Data in Second Language Acquisition. Amsterdam: John Benjamins, 219–228.

Marx, Nicole (2017): Schreibende mit nichtdeutscher Familiensprache. In: Becker-Mrotzek, Michael / Grabowski, Joachim / Steinhoff, Torsten (Hg.): Forschungshandbuch empirische Schreibdidaktik. Münster: Waxmann, 139–152.

Matias, Júlio C. M. (2017): Schreibprozesse im Kontrast. Dissertation. Berlin: Erich Schmidt.

McEnery, Tony / Brezina, Vaclav / Gablasova, Dana / Banerjee, Jayanti (2019): Corpus Linguistics, Learner Corpora, and SLA: Employing Technology to Analyze Language Use. In: Annual Review of Applied Linguistics, 39, 74–92. DOI: 10.1017/S0267190519000096.

Meisel, Jürgen M. (2011): First and Second Language Acquisition. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511862694.

Meisel, Jürgen M. (2021): Diversity and divergence in bilingual acquisition. In: Zeitschrift für Sprachwissenschaft, 40/1, 65–88. DOI: 10.1515/zfs–2021–2025.

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. DOI: 10.1017/S0272263100004137.

Meißner, Cordula / Lange, Daisy / Fandrych, Christian (2016): Korpusanalyse. In: Caspari, Daniela / Klippel, Friederike / Legutke, Michael / Schramm, Karen (Hg.): Forschungsmethoden in der Fremdsprachendidaktik. Ein Handbuch. Tübingen: Narr Francke Attempto, 306–319.

Meng, Katharina / Rehbein, Jochen (Hg. 2007): Kindliche Kommunikation – einsprachig und mehrsprachig. Münster: Waxmann.

MERLIN-Projekt (2014): Nutzerhandbuch. Unter: www.merlin-platform.eu [11.3.2021].

Merkelbach, Christoph (2020): Einführung zum Themenschwerpunkt – Ziel berufliche Integration: DaF-Unterricht für geflüchtete Jugendliche und junge Erwachsene. In: Zeitschrift für Interkulturellen Fremdsprachenunterricht, 25/1, 949–951. Unter: file:///C:/Users/ms/Downloads/zif–3265-merkelbach.pdf [22.02.2023].

Michel, Marije / Murakami, Akira / Alexopoulou, Theodora / Meurers, Detmar (2019): Effects of task type on morphosyntactic complexity across proficiency. In: Instructed Second Language Acquisition (ISLA), 3/2. DOI: 10.1558/isla.38248.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Middeke, Annegret / Eichstaedt, Annett / Jung, Matthias / Kniffka, Gabriele (Hg.) (2017): Wie schaffen wir das? Beiträge zur sprachlichen Integration geflüchteter Menschen. Göttingen: Universitätsverlag Göttingen.

Montanari, Elke (2018): Diagnostik der Schriftsprachaneignung in unterschiedlichen Altersstufen. In: Grießhaber, Wilhelm / Schmölzer-Eibinger, Sabine / Roll, Heike / Schramm, Karen (Hg.): Schreiben in der Zweitsprache Deutsch. Berlin / Boston: De Gruyter, 183–198. DOI: 10.1515/9783110354577–013.

Muntschick, Elisabeth / Portmann, Annette / Schwendemann, Matthias / Wisniewski, Katrin (2020): DISKO (Deutsch im Studium: Lernerkorpus): Handbuch. Unter: https://home.uni-leipzig.de/sprastu/Muntschick_et_al_2020_DISKO-Handbuch.pdf [23.2.2023].

Murakami, Akira (2014): Individual variation and the role of L1 in the L2 development of English grammatical morphemes: insights from learner corpora. Dissertation. Cambridge. DOI: 10.17863/CAM.16509.

Nicholas, Howard / Lenzing, Anke / Roos, Jana (2019): How does PT's view of acquisition relate to the challenge of widening perspectives on SLA? In: Lenzing, Anke / Nicholas, Howard / Roos, Jana (Hg.): Widening Contexts for Processability Theory. Amsterdam: John Benjamins, 391–398. DOI: 10.1075/palart.7.17nic.

Niederhaus, Constanze (2018): Schreiben in der beruflichen Bildung in der Zweitsprache Deutsch. In: Grießhaber, Wilhelm / Schmölzer-Eibinger, Sabine / Roll, Heike / Schramm, Karen (Hg.): Schreiben in der Zweitsprache Deutsch. Berlin / Boston: De Gruyter, 149–164. DOI: 10.1515/9783110354577–011.

Norris, John / Ortega, Lourdes (2009): Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. In: Applied Linguistics, 30/4, 555–578. DOI: 10.1093/applin/amp044.

Oldfield, Richard C. (1971): The assessment and analysis of handedness: The Edinburgh inventory. In: Neuropsychologia, 9/1, 97–113. DOI: 10.1016/0028–3932(71)90067–4.

Opitz, Bertram / Friederici, Angela D. (2004): Brain correlates of language learning: the neuronal dissociation of rule-based versus similarity-based learning. In: The Journal of Neuroscience: the official journal of the Society for Neuroscience, 24/39, 8436–8440. DOI: 10.1523/JNEUROSCI.2220–04.2004.

Ortega, Lourdes (2009): Understanding second language acquisition. London / New York: Routledge.

Ortega, Lourdes (2011): New trends in SLA research: Theories, methods, ethics. Invited lecture at National Tsing Hua University, Taiwan, June 8. [Präsentation]. Unter: https://sites.google.com/a/georgetown.edu/lourdes-ortega/ [23.2.2023].

Ortega, Lourdes (2014): Trying out theories on interlanguage: Description and explanation over 40 years of L2 negation research. In: Han, ZhaoHong / Tarone,

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Elaine (Hg.): Interlanguage: 40 Years Later. Amsterdam: John Benjamins, 173–202. DOI: 10.1075/lllt.39.10ch8.
Ortega, Lourdes (2015): Syntactic complexity in L2 writing: Progress and expansion. In: Journal of Second Language Writing, 29, 82–94. DOI: 10.1016/j.jslw.2015.06.008.
Ortega, Lourdes / Iberri-Shea, Gina (2005): Longitudinal Research In Second Language Acquisition. Recent Trends And Future Directions. In: Annual Review Applied Linguistics 25, 26–45. DOI: 10.1017/S0267190505000024.
Ortega, Lourdes / Byrnes, Heidi (Hg.) (2010): The longitudinal study of advanced L2 capacities. New York: Routledge.
Ortega, Lourdes / Han, ZhaoHong (Hg.) (2017): Complexity Theory and Language Development. In Celebration of Diane Larsen-Freeman. Amsterdam: John Benjamins. DOI: 10.1075/lllt.48.
Orth-Chambah, Jutta / Perlmann, Balme, Michaela / Schwalb, Susanne (2012): Sicher! B1+. Arbeitsbuch. Ismaning: Hueber.
Özkan, Ibrahim / Althoff, Fabian (2019): Fremd in der Fremde. In: Wege zum Menschen, 71/1, 51–59. DOI: 10.13109/weme.2019.71.1.51.
Pallotti, Gabriele (2007): An Operational Definition of the Emergence Criterion. In: Applied Linguistics, 28/3, 361–382. DOI: 10.1093/applin/amm018.
Pallotti, Gabriele (2009): CAF: Defining, Refining and Differentiating Constructs. In: Applied Linguistics, 30/4, 590–601. DOI: 10.1093/applin/amp045.
Pallotti, Gabriele (2015): A simple view of linguistic complexity. In: Second Language Research, 31/1, 117–134. DOI: 10.1177/0267658314536435.
Pallotti, Gabriele (2021): Cratylus' silence: On the philosophy and methodology of Complex Dynamic Systems Theory in SLA. In: Second Language Research. DOI: 10.1177/0267658321992451.
Papi, Mostafa / Teimouri, Yasser (2014): Language Learner Motivational Types: A Cluster Analysis Study. In: Language Learning, 64/3, 493–525. DOI: 10.1111/lang.12065.
Perdue, Clice (Hg. 1993): Adult Language Acquisition: Cross-Linguistic Perspectives. Vol. I: Field Methods. Cambridge: Cambridge University Press.
Perlmann-Balme, Michaela / Schwalb, Susanne (2012): Sicher! B1+. Kursbuch. Ismaning: Hueber.
Perlmann-Balme, Michaela / Schwalb, Susanne (2014): Sicher! B2. Kursbuch. München: Hueber.
Perlmann-Balme, Michaela / Schwalb, Susanne / Matussek, Magdalen (2014): Sicher! B2. Arbeitsbuch. München: Hueber.
Perlmann-Balme, Michaela / Schwalb, Susanne / Matussek, Magdalen (2016): Sicher C1. Kursbuch. München: Hueber.
Perlmann-Balme, Michaela / Schwalb, Susanne / Matussek, Magdalen (2016): Sicher C1. Arbeitsbuch. München: Hueber.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Perry, Bruce D. (2006): Fear and learning: Trauma-related factors in the adult education process. In: New Directions for Adult and Continuing Education, 110, 21–27. DOI: 10.1002/ace.215.

Petersen, Inger (2014): Schreibfähigkeit und Mehrsprachigkeit. Berlin: De Gruyter. DOI: 10.1515/9783110318333.

Petersen, Inger (2019): Messung, Beurteilung und Förderung von Schreibkompetenz in Deutsch als Erst- und Zweitsprache – ein Überblick. In: Kaplan, Ina / Petersen, Inger (Hg.): Schreibkompetenzen messen, beurteilen und fördern. Münster / New York: Waxmann, 11–36.

Pichette, François / Béland, Sébastien / Jolani, Shahab / Leśniewska, Justyna (2015): The handling of missing binary data in language research. In: Studies in Second Language Learning and Teaching, 5/1, 153–169. DOI: 10.14746/ssllt.2015.5.1.8.

Pienemann, Manfred (1984): Psychological Constraints on the Teachability of Languages. In: Studies in Second Language Acquisition, 6/2, 186–214. DOI: 10.1017/S0272263100005015.

Pienemann, Manfred (1998): Language Processing and Second Language Development: Processability Theory. Amsterdam: John Benjamins. DOI: 10.1075/sibil.15.

Pienemann, Manfred (1998a): Developmental dynamics in L1 and L2 acquisition: Processability Theory and generative entrenchment. In: Bilingualism, 1/1, 1–20. DOI: 10.1017/S1366728998000017.

Pienemann, Manfred (Hg.) (2005): Cross-Linguistic Aspects of Processability Theory. Amsterdam: John Benjamins. DOI: 10.1075/sibil.30.

Pienemann, Manfred (2005a): An introduction to Processability Theory. In: Pienemann, Manfred (Hg.): Cross-Linguistic Aspects of Processability Theory. Amsterdam: John Benjamins, 1–60. DOI: 10.1075/sibil.30.03pie.

Pienemann, Manfred (2007): Variation and dynamic systems in SLA. In: Bilingualism, 10/1, 43–46. DOI: 10.1017/S1366728906002793.

Pienemann, Manfred / Johnston, Malcolm (1996): A brief history of processing approaches to SLA: reply to Mello. In: Second Language Research, 12/3, 319–334. DOI: 10.1177/026765839601200304.

Pienemann, Manfred / Di Biase, Bruno / Kawaguchi, Satomi / Håkansson, Gisela (2005a): Processability, typological distance and L1 transfer. In: Pienemann, Manfred (Hg.): Cross-Linguistic Aspects of Processability Theory. Amsterdam: John Benjamins, 85–116. DOI: 10.1075/sibil.30.05pie.

Pienemann, Manfred / Di Biase, Bruno / Kawaguchi, Satomi (2005b): Extending Processability Theory. In: Pienemann, Manfred (Hg.): Cross-Linguistic Aspects of Processability Theory. Amsterdam: John Benjamins, 199–251. DOI: 10.1075/sibil.30.09pie.

Pienemann, Manfred / Keßler, Jörg-Uwe (Hg.) (2011): Studying Processability Theory. An Introductory Textbook. Amsterdam: John Benjamins. DOI: doi.org/10.1075/palart.1.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Pienemann, Manfred / Lenzing, Anke (2015): Processability Theory. In: VanPatten, Bill / Williams, Jessica (Hg.): Theories in second language acquisition. An introduction. 2. Auflage. New York / London: Routledge, 159–179.

Ping, Li / Legault, Jennifer / Litcofsky, Kaitlyn A. (2014): Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. In: Cortex, 58, 301–324. DOI: 10.1016/j.cortex.2014.05.001.

Piniel, Katalin / Csizér, Kata (2015): Changes in Motivation, Anxiety and Self-efficacy during the Course of an Academic Writing Seminar. In: Dörnyei, Zoltán / MacIntyre, Peter D. / Henry, Alastair (Hg.): Motivational dynamics in language learning. Bristol: Multilingual Matters, 164–194. DOI: 10.21832/9781783092574–015.

Piotrowski, Rajmund G. / Altmann, Gabriel / Köhler, Reinhard (Hg.) (2005): Quantitative Linguistik. Berlin / Boston: De Gruyter. DOI: 10.1515/9783110155785.

Plonsky, Luke (2014): Study Quality in Quantitative L2 Research (1990–2010): A Methodological Synthesis and Call for Reform. In: Modern Language Journal, 98/1, 450–470. DOI: 10.1111/j.1540–4781.2014.12058.x.

Plonsky, Luke / Oswald, Frederick L. (2014): How Big Is "Big"? Interpreting Effect Sizes in L2 Research. In: Language Learning, 64/4, 878–912. DOI: 10.1111/lang.12079.

Portmann, Annette (2020): Übernahme von sprachlichen Strukturen aus der schriftlichen Vorgabe in Schreibaufgaben. Eine Lernerkorpusstudie. Unveröffentlichte Masterarbeit. Leipzig.

Rafferty, Anna / Manning, Christopher D. (2008): Parsing Three German Treebanks: Lexicalized and Unlexicalized Baselines. In: ACL Workshop on Parsing German. Unter: https://nlp.stanford.edu/pubs/german-acl08.pdf [23.02.2023].

Rees, Jonathan / Klapper, John (2008): Issues in the quantitative longitudinal measurement of second language progress in the study abroad context. In: Ortega, Lourdes / Byrnes, Heidi (Hg.): The longitudinal study of advanced L2 capacities. New York: Routledge, 89–105.

Reich, Hans H. / Roth, Hans-Joachim / Gantefort, Christoph (2008): Auswertungshinweise ‚Der Sturz ins Tulpenbeet' (Deutsch) und Auswertungsbogen. In: Klinger, Thorsten / Schwippert, Knut / Leiblein, Birgit (Hg.): Evaluation im Modellprogramm FörMig: Planung und Realisierung eines Evaluationskonzepts. Münster / New York: Waxmann, 209–237.

Reich, Hans H. / Roth, Hans-Joachim / Döll, Marion (2009): Auswertungshinweise ‚Fast Catch Bumerang' (Deutsch) und Schreibaufgabe ‚Fast Catch Bumerang'. In: Lengyel, Drorit / Reich, Hans H. / Roth, Hans-Joachim / Döll, Marion (Hg.): Von der Sprachdiagnose zur Sprachförderung. Münster / New York: Waxmann, 209–241

Reuschenbach, Daniel (2009): Nichtparametrische Testverfahren. In: Albers, Sönke / Klapper, Daniel / Konradt, Udo / Walter, Achim / Wolf, Joachim (Hg.): Methodik der empirischen Forschung. Wiesbaden: Gabler, 501–520.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Ricart Brede, Julia (2020): Lernersprachliche Texte im Biologieunterricht. Berlin / Boston: De Gruyter. DOI: 10.1515/9783110687002.

Riedner, Renate / Kassem, Nabil (2010): Kontrastive Analyse Arabisch-Deutsch. In: Krumm, Hans-Jürgen / Fandrych, Christian / Hufeisen, Britta / Riemer, Claudia: Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch. 1. Halbband. Berlin: De Gruyter, 531–538. DOI: 10.1515/9783110240245.518.

Riehl, Claudia Maria / Schroeder, Christoph (2022): DaF/DaZ im Kontext von Mehrsprachigkeit. In: Deutsch als Fremdsprache, 59/2, 67–76. DOI: 10.37307/j.2198–2430.2022.02.02.

Roche, Jörg / Terrasi-Haufe, Elisabetta (2016): Deutschunterricht mit Flüchtlingen. Grundlagen und Konzepte. Tübingen: Narr Francke Attempto.

Roll, Heike (2021): Sprachliche Teilkompetenzen (2): Schriftlichkeit. In: Altmayer, Claus / Biebighäuser, Katrin / Haberzettl, Stefanie / Heine, Antje (Hg.): Handbuch Deutsch als Fremd- und Zweitsprache. Stuttgart: J.B. Metzler, 268–281. DOI: 10.1007/978–3–476–04858–5_17.

Rosa, Elena / Leow, Ronald (2004): Awareness, different learning conditions, and L2 development. In: Applied Psycholinguistics, 25/2, 269–292. DOI: 10.1017/S0142716404001134.

Rosmawati, Rosmawati (2020): Profiling the dynamic changes of syntactic complexity in L2 academic writing. In: Fogal, Gary / Verspoor, Marjolijn (Hg.): Complex Dynamic Systems Theory and L2 Writing Development. Amsterdam: John Benjamins, 109–131. DOI: 10.1075/lllt.54.05ros.

Rösler, Dietmar (2020): Das Verhältnis von gesteuertem und ungesteuertem Zweit- und Fremdsprachenlernen. In: ide. informationen zur deutschdidaktik. Zeitschrift für den Deutschunterricht in Wissenschaft und Schule, 44/4, 10–19.

Röß, Ann-Katrin (2019): Zum Zusammenhang zwischen sozialer Integration und dem Erwerb des Deutschen als Zweitsprache von jungen Erwachsenen mit der Muttersprache Arabisch. Eine quantitative Analyse. Unveröffentlichte Masterarbeit. Leipzig.

Sächsisches Bildungsinstitut (Hg.) (2009): Niveaubeschreibungen Deutsch als Zweitsprache für die Sekundarstufe I. Radebeul: Sächsisches Bildungsinstitut, 11–15.

Schellhardt, Christin / Schroeder, Christoph (2015): Nominalphrasen in deutschen und türkischen Texten mehrsprachiger SchülerInnen. In: Köpcke, Klaus-Michael (Hg.): Deutsche Grammatik in Kontakt. Deutsch als Zweitsprache in Schule und Unterricht. Berlin: De Gruyter, 241–261. DOI: 10.1515/9783110367171–011.

Schiller, Anne / Teufel, Simone / Stöckert, Christine / Thielen, Christine (1999): Guidelines für das Tagging deutscher Textcorpora mit STTS. Technischer Bericht. Institut für maschinelle Sprachverarbeitung, Stuttgart. Unter: http://www.sfs.uni-tuebingen.de/resources/stts–1999.pdf [23.02.2023].

Schleicher, Nora (2018): Die *Elicited Imitation* als Instrument zur Erfassung grammatischer Kompetenz. Unveröffentlichte Masterarbeit. Leipzig.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Schmidt, Thomas / Schütte, Wilfried / Winterscheid, Jenny (2015): cGAT. Konventionen für das computergestützte Transkribieren in Anlehnung an das Gesprächsanalytische Transkriptionssystem 2 (GAT2). Unter: https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/4616/file/Schmidt_Schuette_Winterscheid_cGAT_2015.pdf [23.02.2023].

Schmölzer-Eibinger, Sabine (2010): Textkompetenz. In: Krumm, Hans-Jürgen / Fandrych, Christian / Hufeisen, Britta / Riemer, Claudia: Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch. 2. Halbband. Berlin: De Gruyter, 1130–1137. DOI: 10.1515/9783110240252.1073.

Schmölzer-Eibinger, Sabine (2018): Literalität und Schreiben in der Zweitsprache. In: Grießhaber, Wilhelm / Schmölzer-Eibinger, Sabine / Roll, Heike / Schramm, Karen (Hg.): Schreiben in der Zweitsprache Deutsch. Berlin / Boston: De Gruyter, 3–16. DOI: 10.1515/9783110354577–001.

Schramm, Karen (2018): Methoden zur empirischen Erforschung des Schreibens in Deutsch als Zweitsprache. In: Grießhaber, Wilhelm / Schmölzer-Eibinger, Sabine / Roll, Heike / Schramm, Karen (Hg.): Schreiben in der Zweitsprache Deutsch. Berlin / Boston: De Gruyter, 33–46. DOI: 10.1515/9783110354577–003.

Schwald, Oliver (2017): Spracherwerb trotz Trauma? Wie Traumatisierungen und das Leben als Asylsuchende das Lernen beeinflussen. In: Babylonia, 26/1, 28–31.

Sefgalowitz, Normann (2003): Automaticity and second languages. In: Doughty, Catherine J. / Long, Michael (Hg.): The handbook of second language acquisition. Oxford: Blackwell, 382–408.

Selinker, Larry (1972): Interlanguage. In: IRAL – International Review of Applied Linguistics in Language Teaching, 10/1–4, 209–231. DOI: 10.1515/iral.1972.10.1–4.209.

Selinker, Larry (2014): Interlanguage 40 years on: Three themes from here. In: Han, ZhaoHong / Tarone, Elaine (Hg.): Interlanguage: 40 Years later. Amsterdam: John Benjamins, 221–246. DOI: 10.1075/lllt.39.12ch1.

Selting, Margret et al. (2009): Gesprächsanalytisches Transkriptionssytem 2 (GAT 2). In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion, 10, 353–402. Unter: http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf [23.02.2023].

Settinieri, Julia / Demirkaya, Sevilen / Feldmeier, Alexis / Gültekin-Karakoç, Nazan / Riemer, Claudia (Hg.) (2014): Empirische Forschungsmethoden für Deutsch als Fremd- und Zweitsprache. Eine Einführung. Paderborn: Ferdinand Schöningh.

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. DOI: 10.3389/fpsyg.2021.716485.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Skehan, Peter (1989): Individual differences in second language learning. London: Edward Arnold.

Skehan, Peter (1996): Second language acquisition and task-based instruction. In: Willis, Jane / Willis Dave (Hg.): Challenge and change in language teaching. Oxford: Heinemann, 17–30.

Skehan, Peter (1998): A cognitive approach to language learning. Oxford: Oxford University.

Steinhoff, Torsten (2017): Untersuchung von Textkorpora. In: Becker-Mrotzek, Michael / Grabowski, Joachim / Steinhoff, Torsten (Hg.): Forschungshandbuch empirische Schreibdidaktik. Münster: Waxmann, 353–368.

Spoelman, Marianne / Verspoor, Marjolijn (2010): Dynamic Patterns in Development of Accuracy and Complexity: A Longitudinal Case Study in the Acquisition of Finnish. In: Applied Linguistics, 31/4, 532–553. DOI: 10.1093/applin/amq001.

Steenbeek, Henderien / Jansen, Louise / van Geert, Paul (2012): Scaffolding dynamics and the emergence of problematic learning trajectories. In: Learning and Individual Differences, 22/1, 64–75. DOI: 10.1016/j.lindif.2011.11.014.

Søndergaard, Hans-Peter (2017): Language learning in refugees, an account of a Swedish study. In: Babylonia, 26/1, 15–18.

Taubert, Marco / Draganski, Bogdan / Anwander, Alfred / Müller, Karsten / Horstmann, Anette / Villringer, Arno / Ragert, Patrick (2010): Dynamic properties of human brain structure: learning-related changes in cortical areas and associated fiber connections. In: The Journal of Neuroscience, 30/35, 11670–11677. DOI: 10.1523/JNEUROSCI.2567–10.2010.

Taubert, Marco / Villringer, Arno / Ragert, Patrick (2012): Learning-related gray and white matter changes in humans: an update. In: Neuroscientist, 18/4, 320–325. DOI: 10.1177/1073858411419048.

Taylor, Wayne (2000): Change-Point Analyzer 2.3 shareware program. Taylor Enterprises, Libertyville, Illinois. Unter: http://www.variation.com/cpa [23.02.2023].

Taylor, Wayne (2000|2018): Change-Point Analysis. A Powerful New Tool For Detecting Changes. Unter: https://variation.com/wp-content/uploads/change-point-analyzer/change-point-analysis-a-powerful-new-tool-for-detecting-changes.pdf [23.02.2023].

Thielmann, Winfried (2021): Mozart für Anfänger? Zum Verhältnis von Sprachdidaktik, Spracherwerbsforschung und Gemeinsamem Europäischem Referenzrahmen. In: Informationen Deutsch als Fremdsprache, 48/6, 648–670. DOI: 10.1515/infodaf–2021–0083.

Thurmair, Maria / Willkop, Eva-Maria (Hg. 2003): Am Anfang war der Text. 10 Jahre „Textgrammatik der deutschen Sprache". München: iudicium.

Tracy, Rosemarie (2007): Wie Kinder Sprachen lernen. Und wie wir sie dabei unterstützen können. Tübingen: Narr Francke Attempto.

Tracy, Rosemarie (2007a): Wieviele Sprachen passen in einen Kopf? Mehrsprachigkeit als Herausforderung für Gesellschaft und Forschung. In: Anstatt,

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Tanja (Hg.): Mehrsprachigkeit bei Kindern und Erwachsenen. Erwerb, Formen, Förderung. Tübingen: Narr Francke Attempto, 69–92.

Tschirner, Erwin (2015): Assessing evidence of validity and reliability of the ACTFL Reading Proficiency Test (RPT) (Report No. 2015/3-PUB–1). Leipzig: Institut für Testforschung und Testentwicklung.

Tschirner, Erwin (2015a): Assessing evidence of validity and reliability of the ACTFL Listening Proficiency Test (LPT) (Report No. 2015/3-PUB–2). Leipzig: Institut für Testforschung und Testentwicklung.

Tschirner, Erwin (2016): Listening and Reading Proficiency Levels of College Students. In: Foreign Language Annals, 49/2, 201–223. DOI: 10.1111/flan.12198.

Tschirner, Erwin (2020): Assessing evidence of validity and reliability of the ACTFL Reading Proficiency Test (RPT) (Report No. 2020/2-PUB–2). Leipzig: Institut für Testforschung und Testentwicklung.

Tschirner, Erwin (2020a): Assessing evidence of validity and reliability of the ACTFL Listening Proficiency Test (LPT) (Report No. 2020/2-PUB–3). Leipzig: Institut für Testforschung und Testentwicklung.

Tschirner, Erwin / Bärenfänger, Olaf / Wisniewski, Katrin (2015): Assessing Evidence of Validity of the ACTFL CEFR Listening and Reading Proficiency Tests (LPT and RPT) Using a Standard-Setting Approach (Technical Report 2015-EU-PUB–2). Leipzig: Leipzig: Institut für Testforschung und Testentwicklung.

Ullmann, Michael (2016): The Declarative/Procedural Model: A Neurobiological Model of Language Learning, Knowledge, and Use. In: Neurobiology of Language, 953–968. DOI: 10.1016/B978–0–12–407794–2.00076–6.

Ullman, Michael / Lovelett, Jarrett (2018): Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. In: Second Language Research, 34/1, 39–65. DOI: 10.1177/0267658316675195.

Ulrich, Tanja (2017): Grammatikerwerb und grammatische Störungen im Kindesalter. Ergebnisse des Forschungsprojekts GED 4–9 und ihre Implikationen für sprachdiagnostische und -therapeutische Methoden. Habilitationsschrift. Universität Köln. Unter: https://kups.ub.uni-koeln.de/9011/1/Habilitation_Ulrich_OnlinePub.pdf [23.02.2023].

van Dijk, Marijn / van Geert, Paul (2007): Wobbles, humps and sudden jumps: a case study of continuity, discontinuity and variability in early language development. In: Infant and Child Development, 16/1, 7–33. DOI: 10.1002/icd.506.

van Dijk, Marijn / Verspoor, Marjolijn / Lowie, Wander (2011): Variability and DST. In: Verspoor, Marjolijn / de Bot, Kees / Lowie, Wander (Hg.): A dynamic approach to second language development. Methods and techniques. Amsterdam, Philadelphia: John Benjamins, 55–84. DOI: 10.1075/lllt.29.04van.

van Geert, Paul (2008): The Dynamic Systems Approach in the Study of L1 and L2 Acquisition: An Introduction. In: Modern Language Journal, 92/2, 179–199. DOI: 10.1111/j.1540–4781.2008.00713.x.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

van Moere, Alistair (2012): A psycholinguistic approach to oral language assessment. In: Language Testing, 29, 325–344. DOI: 0.1177/0265532211424478.
VanPatten, Bill (2014): The limits of instruction: 40 years after "Interlanguage". In: Han, ZhaoHong / Tarone, Elaine (Hg.): Interlanguage: 40 Years later. Amsterdam: John Benjamins, 105–126. DOI: 10.1075/lllt.39.07ch5.
VanPatten, Bill / Williams, Jessica (Hg. 2015): Theories in second language acquisition: An introduction. 2. Auflage. New York: Routledge.
VanPatten, Bill / Williams, Jessica (2015a): Early Theories in SLA. In: VanPatten, Bill / Williams, Jessica (Hg.): Theories in second language acquisition. An introduction. 2. Auflage. New York / London: Routledge, 17–33.
Verspoor, Marjolijn / Lowie, Wander / van Dijk, Marijn (2008): Variability in Second Language Development From a Dynamic Systems Perspective. In: Modern Language Journal, 92/2, 214–231. DOI: 10.1111/j.1540–4781.2008.00715.x.
Verspoor, Marjolijn / de Bot, Kees / Lowie, Wander (Hg.) (2011): A dynamic approach to second language development. Methods and techniques. Amsterdam, Philadelphia: John Benjamins. DOI: 10.1075/lllt.29.
Verspoor, Marjolijn / van Dijk, Marijn (2011): Visualizing Interaction between variables. In: Verspoor, Marjolijn / de Bot, Kees / Lowie, Wander (Hg.): A dynamic approach to second language development. Methods and techniques. Amsterdam, Philadelphia: John Benjamins, 85–98. DOI: 10.1075/lllt.29.05ver.
Verspoor, Marjolijn / Schmid, Monika S. / Xu, Xiaoyan (2012): A dynamic usage based perspective on L2 writing. In: Journal of Second Language Writing, 21/3, 239–263. DOI: 10.1016/j.jslw.2012.03.007.
Vinther, Thora (2002): Elicited Imitation: a brief overview. In: International Journal of Applied Linguistics, 12/1, 54–73. DOI: 10.1111/1473–4192.00024.
Vyatkina, Nina (2012): The Development of Second Language Writing Complexity in Groups and Individuals: A Longitudinal Learner Corpus Study. In: The Modern Language Journal, 96/4, 576–598. DOI: 10.1111/j.1540–4781.2012.01401.x.
Vyatkina, Nina (2013): Specific Syntactic Complexity: Developmental Profiling of Individuals Based on an Annotated Learner Corpus. In: Modern Language Journal, 97/1, 11–30. DOI: 10.1111/j.1540–4781.2012.01421.x.
Vyatkina, Nina (2016): The Kansas Developmental Learner corpus (KANDEL). In: International Journal for Learner Corpus Research, 2/1, 101–119. DOI: 10.1075/ijlcr.2.1.04vya.
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. DOI: 10.1016/j.jslw.2015.06.006.
Walter, Maik / Doolittle Seanna (2012): Vorfeldfortschritte im Deutschen als Fremdsprache. Eine lernerkorpusbasierte Longitudinalanalyse amerikanischer Deutschlerner. In: Deutsch als Fremdsprache, 49/1, 3–11. DOI: 10.37307/j.2198–2430.2012.01.02.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Wegener, Heide (1988–1992): Augsburger Korpus. In: The Language Archive.
Wegener, Heide (1995): Die Nominalflexion des Deutschen – verstanden als Lerngegenstand. Tübingen: Niemeyer. Zugleich Teildruck von: Wegener, Heide (1993): Kindlicher Zweitspracherwerb. Universität Augsburg. Habilitationsschrift.
Wei, Xuehu / Adamson, Helyne / Schwendemann, Matthias / Goucha, Tomás / Friederici, Angela D. / Anwander, Alfred (2023): Nativ language differences in the structural connectome of the human brain. In: NeuroImage. DOI: 10.1016/j.neuroimage.2023.119955.
Weinrich, Harald (2007): Textgrammatik der deutschen Sprache. Hildesheim: Georg Olms.
Weinzierl, Christian / Wrobel, Arne (2017): Schreibprozesse untersuchen. In: Becker-Mrotzek, Michael / Grabowski, Joachim / Steinhoff, Torsten (Hg.): Forschungshandbuch empirische Schreibdidaktik. Münster: Waxmann, 221–237.
Wildgen, Wolfgang / Plath, Peter Jörg (2005): Katastrophen- und Chaostheorie in der linguistischen Modellbildung. In: Piotrowski, Rajmund G. / Altmann, Gabriel / Köhler, Reinhard (Hg.): Quantitative Linguistik. Boston / Berlin: De Gruyter, 688–705. DOI: 10.1515/9783110155785.
Willis, Jane / Willis Dave (Hg.) (1996): Challenge and change in language teaching. Oxford: Heinemann.
Winkler, Steffi (2021): New perspectives on teachability. The acquisition of German (S)OV word order in classroom contexts. Dissertation. Münster. Unter: https://nbn-resolving.de/urn:nbn:de:hbz:6–68069404445 [23.02.2023].
Wisniewski, Katrin (2014): Die Validität der Skalen des Gemeinsamen europäischen Referenzrahmens für Sprachen. Frankfurt am Main: Peter Lang. DOI: 10.3726/978–3–653–03925–2.
Wisniewski, Katrin (2018): Sprache und Studienerfolg von Bildungsausländerinnen und -ausländern: Eine Längsschnittstudie an den Universitäten Leipzig und Würzburg. In: Informationen Deutsch als Fremdsprache, 45/4, 573–597. DOI: 10.1515/infodaf–2018–0074.
Wisniewski, Katrin (2020): SLA developmental stages in the CEFR-related learner corpus MERLIN. In: International Journal for Learner Corpus Research, 6/1, 1–37. DOI: 10.1075/ijlcr.18008.wis.
Wisniewski, Katrin / Möhring, Jupp (2021): Die Vergleichbarkeit von Sprachprüfungen zum Hochschulzugang im Licht ausgewählter *post entry*-Sprachtests. In: Deutsch als Fremdsprache, 58/4, 195–213. DOI: 10.37307/j.2198–2430.2021.04.03.
Wisniewski, Katrin / Lenhard, Wolfgang / Spiegel, Leonore / Möhring, Jupp (Hg.) (2022.): Sprache und Studienerfolg bei Bildungsausländerinnen und Bildungsausländern. Münster: Waxmann.
Wisniewski, Katrin / Muntschick, Elisabeth / Portmann, Annette (2022a): Schreiben in der Studiersprache Deutsch: Das Lernerkorpus DISKO. In: Wisniewski, Katrin / Lenhard, Wolfgang / Spiegel, Leonore / Möhring, Jupp (Hg.): Sprache

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

und Studienerfolg bei Bildungsausländerinnen und Bildungsausländern. Münster: Waxmann, 283–304.

Wode, Henning (1993): Psycholinguistik. Eine Einführung in die Lehr- und Lernbarkeit von Sprachen. Ismaning: Hueber.

Wolfe-Quintero, Kate / Inagaki, Shunji / Kim, Hae-Young (1998): Second language development in writing. Measures of fluency, accuracy, & complexity. Honolulu Hawaii: Univ. of Hawai'i Press (Technical report / Second Language Teaching & Curriculum Center, 17).

Wolff, Armin / Winters-Ohle, Elmar (Hg.) (2001): Wie schwer ist die deutsche Sprache wirklich? Regensburg: Fachverband Deutsch als Fremdsprache.

Wöllstein, Angelika (2010): Topologisches Satzmodell. Heidelberg: Universitätsverlag Winter.

Wöllstein, Angelika / Dudenredaktion (Hg.) (2016): Duden. Die Grammatik. Unentbehrlich für richtiges Deutsch. 9., vollständige überarbeitete und aktualisierte Auflage. Berlin: Bibliographisches Institut.

Zabel, Rebecca / Wisniewski. Katrin / Ketzer-Nöltge, Almut / Nestler, Doreen (2022): DaF und DaZ: Schnittstellen, Divergenzen, Potenziale. Einführung in den Themenschwerpunkt. In: Deutsch als Fremdsprache, 59/1, 3–14.

Ziegler, Arne / Köpcke, Klaus-Michael (Hg.) (2015): Deutsche Grammatik in Kontakt. Deutsch als Zweitsprache in Schule und Unterricht. Berlin / Boston: De Gruyter. DOI: 10.1515/9783110367171.

Zifonun, Gisela / Hoffmann, Ludger / Strecker, Bruno (1997): Grammatik der deutschen Sprache. Berlin: De Gruyter.

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

7 Tabellenverzeichnis

Tabelle 1: Schematische Übersicht des topologischen Satzmodells 15
Tabelle 2: Übersicht über verschiedene Modelle syntaktischer Entwicklungsstufen 62
Tabelle 3: Zugangsvoraussetzungen und eingesetzte Testverfahren 106
Tabelle 4: Übersicht über den zeitlichen Aufbau der Studie 115
Tabelle 5: Vergleich zwischen Dropouts und Teilnehmenden 127
Tabelle 6: Gruppeneinteilung in Bezug auf die experimentelle Manipulation 132
Tabelle 7: Teilnehmende innerhalb der einzelnen Kursgruppen 133
Tabelle 8: Vergleich zwischen ausgeschlossenen Teilnehmenden und Teilstichprobe 134
Tabelle 9: Übersicht über die durchgeführten Sprachtests 135
Tabelle 10: Überblick über die Ergebnisse der schriftlichen Teile der Probeprüfung und des Goethe-Zertifikat B1 139
Tabelle 11: Profilstufen nach Grießhaber (2012) 145
Tabelle 12: Korrelationen von MLC, DC/C und ComNP/C. MLC 154
Tabelle 13: Hintergrundvariablen TN024 und TN025 160
Tabelle 14: Ausgewählte Texte für TN024 und TN025 161
Tabelle 15: Deskriptive Daten zu den produzierten Texten zu Zeitpunkt 1 (TP1) 172
Tabelle 16: Deskriptive Daten zu den produzierten Texten zu Zeitpunkt 2 (TP2) 174
Tabelle 17: Deskriptive Daten zu den produzierten Texten zu Zeitpunkt 3 (TP3) 175
Tabelle 18: Deskriptive Statistiken für die zu TP1 produzierten Wortstellungsmuster 177
Tabelle 19: Häufigkeitsübersicht über die zu TP1 erreichten Profilstufen 178
Tabelle 20: Deskriptive Statistiken für die zu TP2 produzierten Wortstellungsmuster 179
Tabelle 21: Häufigkeitsübersicht über die zu TP2 erreichten Profilstufen 179
Tabelle 22: Deskriptive Statistiken für die zu TP3 produzierten Wortstellungsmuster 181
Tabelle 23: Häufigkeitsübersicht über die zu TP3 erreichten Profilstufen 181
Tabelle 24: Deskriptive Daten zum Friedman-Test 183
Tabelle 25: Zusammenfassung der Ergebnisse für den Friedman-Test 184
Tabelle 26: Vergleich der beiden Emergenzkriterien 188
Tabelle 27: Deskriptive Statistiken der einfaktoriellen ANOVA für die 3-Clusterlösung 194
Tabelle 28: Zusammenfassung der einfaktoriellen ANOVA für die 3-Clusterlösung 195

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Tabelle 29: Tukey Post-Hoc-Tests für Mehrfachvergleiche plus Bootstrap 196
Tabelle 30: Anzahl der Fälle in jedem Cluster zu Zeitpunkt 1 (TP1) 197
Tabelle 31: Clusterzentren zu Zeitpunkt 1 (TP1) 197
Tabelle 32: Anzahl der Fälle in jedem Cluster zu Zeitpunkt 2 (TP2) 200
Tabelle 33: Clusterzentren zu Zeitpunkt 2 (TP2) 200
Tabelle 34: Anzahl der Fälle in jedem Cluster zu Zeitpunkt 2 (TP2) 201
Tabelle 35: Clusterzentren zu Zeitpunkt 3 (TP3) 201
Tabelle 36: Verlauf der Clusterzugehörigkeiten über die drei Zeitpunkte 202
Tabelle 37: Verlaufskurve anhand der Clusteranalysen – Gruppenbildung ... 203
Tabelle 38: Deskriptive Statistik zum Vergleich der Instruktionsgruppen: ‚Durchschnittliche Länge einer MSE' 207
Tabelle 39: Zusammenfassung der zweifachen ANOVA mit Messwiederholung: ‚Durchschnittliche Länge einer MSE' 208
Tabelle 40: Post-hoc Vergleich mit Bonferroni-Korrektur zur zweifachen ANOVA mit Messwiederholung: ‚Durchschnittliche Länge einer MSE' 208
Tabelle 41: Deskriptive Statistik zum Vergleich der Instruktionsgruppen: ‚Subordinationsrate pro MSE' 209
Tabelle 42: Zusammenfassung der zweifachen ANOVA mit Messwiederholung: ‚Subordinationsrate pro MSE' 209
Tabelle 43: Vergleich mit Bonferroni-Korrektur zur zweifachen ANOVA mit Messwiederholung: ‚Subordinationsrate pro MSE' 209
Tabelle 44: Deskriptive Statistik zum Vergleich der Instruktionsgruppen: ‚Rate komplexer Nominalphrase pro MSE' 210
Tabelle 45: Zusammenfassung der zweifachen ANOVA mit Messwiederholung: ‚Rate komplexer Nominalphrase pro MSE' 210
Tabelle 46: Vergleich mit Bonferroni-Korrektur zur zweifachen ANOVA mit Messwiederholung: ‚Rate komplexer Nominalphrasen pro MSE' 211
Tabelle 47: Kreuztabelle zum Vergleich der Zugehörigkeit 212
Tabelle 48: Zusammenfassung des Chi Quadrat-Tests 213
Tabelle 49: Deskriptive Daten für TN024 und TN25 (aggregiert) 217
Tabelle 50: Deskriptive Daten für TN024 und TN25 – erste Studienphase bis B1 218
Tabelle 51: Deskriptive Daten für TN024 und TN25 – zweite Studienphase bis B2/C1 218
Tabelle 52: Deskriptive Daten für TN024 – Gesamtstudie 219
Tabelle 53: Deskriptive Daten für TN025 – Gesamtstudie 220
Tabelle 54: Deskriptive Daten für TN024 – erste Studienphase 221
Tabelle 55: Deskriptive Daten zu TN024 – zweite Studienphase 222
Tabelle 56: Deskriptive Daten zu TN025 – erste Studienphase 223
Tabelle 57: Deskriptive Daten zu TN025 – zweite Studienphase 224
Tabelle 58: Übersicht Daten TN024 226
Tabelle 59: Übersicht Daten TN025 227
Tabelle 60: Generelle Korrelationen – Wortstellungsmuster (TN024) 256

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Tabelle 61: Generelle Korrelationen – Wortstellungsmuster (TN025) 257
Tabelle 62: Generelle Korrelationen – Syntaxvariablen (TN024)..... 266
Tabelle 63: Generelle Korrelationen – Syntaxvariablen (TN025)..... 267

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

8 Abbildungsverzeichnis

Abbildung 1: Studienphase 1 von Mai 2016 – Februar 2017 (schematisch) ... 121
Abbildung 2: Studienphase 2 von März 2017 – September 2017 (schematisch) ... 121
Abbildung 3: Altersverteilung zu Beginn der Studie (Teilstichprobe, n=44) ... 131
Abbildung 4: Vergleich Probeprüfung – Goethe-Zertifikat B1 ... 138
Abbildung 5:Schreibaufgabe aus dem A1/A2-Test; im Rahmen der Studie erstellt. ... 140
Abbildung 6:Schreibaufgabe aus dem A2-Test; im Rahmen der Studie erstellt. ... 140
Abbildung 7: Schreibaufgabe a aus Modellsatz für das Goethe-Zertifikat B1 a ... 141
Abbildung 8: Schreibaufgabe b aus Modellsatz für das Goethe-Zertifikat B1. ... 142
Abbildung 9: Beispiel für Transkription und Segmentierung in MSE ... 143
Abbildung 10: Schematischer Ablauf unterschiedlicher Clusteranalysen ... 147
Abbildung 11: Schematischer Ablauf einer longitudinalen Clusteranalyse ... 150
Abbildung 12 Texte zu Zeitpunkten TN024 und TN025 ... 162
Abbildung 13: Produzierte Texte im Zeitverlauf der Studie ... 162
Abbildung 14: CUSUM-Diagramm einer CPA für eine idealisierte Lernerin ... 166
Abbildung 15: Output einer CPA für eine idealisierte Lernerin ... 167
Abbildung 16: Tabellarischer Output einer CPA ... 168
Abbildung 17: Boxplot zu den erreichten Profilstufen zu TP1, TP2 und TP3 ... 182
Abbildung 18: Zuordnungsübersicht zur Bestimmung der Clusteranzahl ... 192
Abbildung 19: Dendrogramm der HCA zur Bestimmung der Clusterzahl ... 193
Abbildung 20: Gruppierte Streudiagramme zu Zeitpunkt 1 ... 198
Abbildung 21: Gruppierte Streudiagramme zu Zeitpunkt 2 ... 198
Abbildung 22: Gruppierte Streudiagramme zu Zeitpunkt 3 ... 199
Abbildung 23: Verlaufskurve anhand der longitudinalen Cluster für die Variable ‚Durchschnittliche Länge einer MSE' ... 203
Abbildung 24: Verlaufskurve anhand der longitudinalen Cluster für die Variable ‚Subordinationsrate pro MSE' ... 204
Abbildung 25: Verlaufskurve anhand der longitudinalen Cluster für die Variable ‚Rate komplexer Nominalphrasen pro MSE' ... 205
Abbildung 26: Mittelwertvergleich der drei Clustervariablen ... 207
Abbildung 27: Longitudinale Cluster nach Zugehörigkeit zur-Experimentalgruppe ... 212
Abbildung 28: min-max Diagramm der erreichten Profilstufen (TN024) ... 229
Abbildung 29: min-max Diagramm der erreichten Profilstufen (TN025) ... 230

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Abbildung 30: min-max Diagramm der durchschnittlichen Länge einer MSE (TN024) 231
Abbildung 31: min-max Diagramm der durchschnittlichen Länge einer MSE (TN025) 232
Abbildung 32: min-max Diagramm der Subordinationsrate (TN024) 233
Abbildung 33: min-max Diagramm der Subordinationsrate (TN025) 234
Abbildung 34: min-max Diagramm zur Rate komplexer Nominalphrasen pro MSE (TN024) 235
Abbildung 35: min-max Diagramm zur Rate komplexer Nominalphrasen pro MSE (TN025) 235
Abbildung 36: Interaktion zwischen den Wortstellungsmustern für TN024 .. 238
Abbildung 37: Durch Moving Average geglättete Interaktion zwischen den produzierten Wortstellungsmustern für TN024 241
Abbildung 38: Durch Moving Average geglättete Interaktion zwischen den produzierten Wortstellungsmustern für TN024 242
Abbildung 39: Interaktion zwischen den Wortstellungsmustern für TN025 .. 243
Abbildung 40: Durch Moving Average geglättete Interaktion zwischen den produzierten Wortstellungsmustern für TN025 245
Abbildung 41: Durch Moving Average geglättete Interaktion zwischen den produzierten Wortstellungsmustern für TN025 246
Abbildung 42: Normalisierte Interaktion der Syntaxvariablen für TN024 248
Abbildung 43: Normalisierter und mit einem Moving Average geglättete Interaktion zwischen den drei Syntaxvariablen für TN024 249
Abbildung 44: Normalisierte Interaktion der Syntaxvariablen für TN025 250
Abbildung 45: Normalisierter und mit einem Moving Average geglättete Interaktion zwischen den drei Syntaxvariablen für TN025 251
Abbildung 46: Gleitende Korrelation ‚Finitum' – ‚Separation' 259
Abbildung 47: Gleitende Korrelation ‚Finitum' – ‚Inversion' 261
Abbildung 48: Gleitende Korrelation ‚Finitum' – ‚Verbendstellung' 261
Abbildung 49: Gleitende Korrelation ‚Separation' – ‚Inversion' 263
Abbildung 50: Gleitende Korrelation ‚Separation' – ‚Verbendstellung' 264
Abbildung 51: Gleitende Korrelation ‚Inversion' – ‚Verbendstellung' 265
Abbildung 52: Gleitende Korrelation ‚MLC_norm' – ‚DC/C_norm' 269
Abbildung 53: Gleitende Korrelation ‚MLC_norm' – ‚ComNP/C_norm' 270
Abbildung 54: Gleitende Korrelation ‚DC/C_norm' – ‚ComNP/C_norm' 271
Abbildung 55: Grafischer Output der CPA der erreichten Profilstufen für TN024 273
Abbildung 56: Grafischer Output der CPA der erreichten Profilstufen für TN025 274
Abbildung 57: Grafischer Output der CPA für die Variable ‚Durchschnittliche Länge MSE' (MLC) für TN024 276
Abbildung 58: Grafischer Output der CPA für die Variable ‚Durchschnittliche Länge MSE' (MLC) für TN025 277

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Abbildung 59: Grafischer Output der CPA für die Variable ‚Subordination' (DC/C) für TN024 278
Abbildung 60: Grafischer Output der CPA für die Variable ‚Subordination' (DC/C) für TN025 279
Abbildung 61: Grafischer Output der CPA für die Variable ‚Komplexe Nominalphrasen' (ComNP/C) für TN024 280
Abbildung 62: Grafischer Output der CPA für die Variable ‚Komplexe Nominalphrasen' (ComNP/C) für TN025 281

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

9 Abkürzungsverzeichnis

Abb.	Abbildung
ANOVA	Analysis of Variance (Varianzanalyse)
BAMF	Bundesamt für Migration und Flüchtlinge
CAF	Complexity, Accuracy, Fluency
CDST	Complex Dynamic Systems Theorie
CN/C	Complex Nominals Ratio
CP	Change Point
CPA	Change Point-Analyse
CT	Complexity Theory
DaF	Deutsch als Fremdsprache
DaZ	Deutsch als Zweitsprache
DC/C	Dependent Clause Ratio
DOI	Digital Object Identifier
DST	Dynamic Systems Theory
fMRT	funktionelle Magnetresonanztomographie
GER	Gemeinsamer europäischer Referenzrahmen für Sprachen
HCA	Hierarchische Clusteranalyse
HI	Herder-Institut der Universität Leipzig
IT	TInstitut für Testforschung und Testentwicklung e.V., Leipzig
LFG	Lexical Functional GrammarW
L1	Erstsprache
L2	Zweitsprache
Mdn	Median
MLC	Mean Length of Clause
MWL	Mean Word Length
MSE	Minimale satzwertig Einheit
MSL	Mean Sentence Length
MPI CBS	Max-Planck-Institut für Kognitions- und Neurowissenschaften, Leipzig
MRT	Magnetresonanztomographie
PA	Profilanalyse (nach Grießhaber)
PT	Processability Theory
SD	Standardabweichung
SLA	Second Language Acquisition
TN00X	Teilnehmende/r X
TP1, TP2, TP3	Zeitpunkt 1, 2, 3
UG	Universalgrammatik

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).

Diese Studie untersucht die Entwicklung syntaktischer Strukturen bei erwachsenen Lernenden des Deutschen als Zweitsprache mit der Erstsprache Arabisch. Hierzu werden schriftliche Daten aus einer longitudinalen Studie verwendet, die über insgesamt 18 Monate am Max-Planck-Institut für Kognitions- und Neurowissenschaften in Kooperation mit dem Herder-Institut der Universität Leipzig durchgeführt wurde.

Die Studie verfolgt zwei methodische Perspektiven: Einerseits werden Daten auf Gruppenebene untersucht. Hier kommen neben Profilanalysen longitudinale Clusteranalysen zum Einsatz. Andererseits nimmt die Studie zwei Lernende in einer vergleichenden Einzelfallanalyse in den Blick. Hier bilden ebenfalls Profilanalysen den Ausgangspunkt der Datenanalyse. Diese werden durch grafische Variabilitäts- und Trendanalysen mit unterschiedlichen Korrelationsanalysen und Change Point-Analysen ergänzt.

Ziel dabei ist es, einerseits zu überprüfen, ob sich sequenzielle und implikationelle Erwerbsfolgen für die Lernerinnen und Lerner herausarbeiten lassen und andererseits die komplex-dynamischen Entwicklungen ausgewählter syntaktischer Strukturen im Zweitspracherwerb erwachsener Lernender detailliert zu modellieren.

www.ESV.info

© Copyright Erich Schmidt Verlag GmbH & Co. KG, Berlin 2024. Open Access. Creative Commons-Lizenz 4.0 (BY-NC-ND).