

10 Forschungsdesign

10.1 Forschungsfrage und Erkenntnisinteresse

Mit der KvDaZ-Studie konnten Auswirkungen einer kernvokabularbasierten Sprachförderung nach dem KvDaZ-Konzept auf die Wortschatz- und Grammatikentwicklung herausgearbeitet werden. Unklar bleibt, welche Kernvokabularwörter und welche festen Wortkombinationen die untersuchten Kinder zum ersten Messzeitpunkt bereits aktiv verwendet haben. Erkenntnisse zum natürlichen mündlichen Sprachgebrauch in der frühen Phase der Aneignung von Deutsch als Zweitsprache können genutzt werden, um wissenschaftlich basierte Empfehlungen zur Vokabularauswahl und -vermittlung im Kontext mehrsprachiger UK-Interventionen bereitzustellen.

Vor diesem Hintergrund und unter Beachtung der im Theorieteil aufgedeckten Bedarfe wird die Herleitung folgender Forschungsfrage begründet:

Welches Kernvokabular und welche festen Wortkombinationen nutzen Kinder in der frühen Phase der Aneignung von Deutsch als Zweitsprache im natürlichen mündlichen Sprachgebrauch häufig?

Zur Beantwortung der Forschungsfrage wurde ein *quantitatives nicht-experimentelles deskriptives Forschungsdesign* gewählt. Folgende Argumente begründen die Entscheidung (auch Döring & Bortz, 2016, S. 201ff.):

- *Nicht-kausale Forschungsfrage:* Die Beantwortung der Forschungsfrage erfordert keine experimentelle Manipulation der abhängigen Variablen (Kernvokabular, feste Wortkombinationen). Das Erkenntnisinteresse lag auf dem natürlichen mündlichen Sprachgebrauch.
- *Heterogenität im Feld:* Die vielfältigen Einflussfaktoren auf den DaZ-Erwerb (z.B. Entwicklungsstand in der Erstsprache, strukturelle Rahmenbedingungen der Beschulung, Kap. 2.5) und damit potentielle Störvariablen konnten nicht systematisch kontrolliert werden.
- *Nähe zum Feld:* Durch ein nicht-experimentelles Design lässt sich eine größtmögliche Nähe zum Feld während der Datenerhebung herstellen, sodass auf Grundlage der Ergebnisse praxisnahe Implikationen abgeleitet werden können.

Ein quantitatives Vorgehen erfordert die Operationalisierung der Untersuchungsgegenstände (Meißner, Lange & Fandrych, 2016, S. 309). Das *Kernvokabular* wurde in Anlehnung an die deutschsprachige Forschung von Boenisch (2014a; 2014b) im Hinblick auf den Umfang operationalisiert. Als Häufigkeitskriterium (cut-off point) wurde die 80 %-Marke gewählt. »Besonders bei Personengruppen mit (noch) sehr kleinem Wortschatz (inkl. Zweitspracherwerb) ermöglicht die 80-Prozent-Marke eine relativ eindeutige Zuordnung des Wortschatzes zum Kern- und Randvokabular« (Boenisch, 2014a, S. 166). Demzufolge lässt sich die 80 %-Marke unabhängig von der Gesamtgröße des untersuchten Wortschatzes anwenden (Boenisch, 2014a, S. 166). Zusätzlich wurde in Anlehnung an den internationalen Diskurs das Streuungskriterium $\geq 50\%$ ($n = 11$) auf die 80 %-Marke angewendet, um ein noch feineres Bild über den Gebrauch des Kernvokabulars innerhalb der Untersuchungsgruppe zu gewinnen.

Der Untersuchungsgegenstand der *festen Wortkombinationen* wurde in der aktuellen Forschung bisher nicht eindeutig operationalisiert. Zahlreiche Verwendungsweisen und Definitionen konnten aufgedeckt werden (Kap. 7.2.1, Tab. 11). Mit der Untersuchung der festen Wortkombinationen wurde primär das Interesse verfolgt, die am häufigsten genutzten Wortkombinationen zu erfassen, ohne die Analyse auf vorab festgelegte morphosyntaktische Kategorien (z.B. Verb + Argumentstruktur) zu reduzieren. Das Konstrukt »lexical bundles« wurde als besonders geeignet eingeschätzt.

»Lexical bundles are extracted automatically from raw data—typically using sizable corpora—disregarding any pre-defined linguistic categories. They typically do not coincide with traditional grammatical units, but instead represent clause or phrase fragments« (Ädel & Erdmann, 2012, S. 82).

Der Terminus ist im Kontext der Computer- und Korpuslinguistik verbreitet (Shadrova, 2013, S. 31). Im Gegensatz zu Kollokationen, die in der Regel nach ihrer Struktur klassifiziert werden, wird bei lexical bundles das Frequenzkriterium herangezogen (Crossley & Salsbury, 2011, S. 3). Zu den lexical bundles zählen die am häufigsten verwendeten Wortkombinationen in einem ausgewählten Register¹ (Biber & Barbieri, 2007, S. 264, 267). In der vorliegenden Arbeit wurde der mündliche Sprachgebrauch in der Schule, einschließlich der verschiedenen Kommunikationsbereiche (z.B. Unterricht, Essen, Pause) als ein Register betrachtet. Biber und Barbieri (2007) bezeichnen lexical bundles als einen ergänzenden Ansatz in der Forschung zu festen Wortkombinationen (S. 264). »Lexical bundles are usually not structurally complete and not idiomatic in meaning, but they serve important discourse functions in both spoken and written texts« (Biber & Barbieri, 2007, S. 264). Auf dem Flexibilitätskontinuum nach Shadrova (2013, Tab. 14) sind lexical bundles am flexiblen Ende zu verorten, »weil sie für die Suche nach verschiedenen wiederkehrenden Einheiten und ihren jeweiligen Varianten gebraucht werden« (S. 32). Für die Identifizierung von festen Wortkombinationen und der Beschreibung des Umfangs wurde in der vorliegenden Untersuchung ein frequenzorientierter Ansatz gewählt (Biber

1 Mit dem Begriff »Register« werden Sprech- oder Schreibweisen assoziiert, die je nach Kommunikationsbereich als angemessen gelten (Lang-Groth, 2019, S. 1). Eine alternative Umschreibung wäre Stilebene (Horstmann et al., 2020, S. 141).

& Barbieri, 2007, S. 267; Shadrova, 2013, S. 31). »The frequency cut-off used to identify lexical bundles is somewhat arbitrary« (Biber & Barbieri, 2007, S. 267). Das Häufigkeitskriterium wurde aufgrund fehlender Referenzwerte *heuristisch* und mithilfe der Analyse der Häufigkeit-Rangbeziehung bestimmt (Verlauf des Liniendiagramms bzw. der Zipf-Kurve, Abstand der Ränge). Die Länge der festen Wortkombinationen wurde angelehnt an verschiedene andere Untersuchungen auf *Dreiwortkombinationen* begrenzt (auch Clendon & Erickson, 2008; Clendon, Sturm & Cali, 2013; Ebeling & Hasselgård, 2015; Stuart et al., 1997; Northbrook, Allen & Conklin, 2021; Simpson-Vlach & Ellis, 2010). Nach Simpson-Vlach und Ellis (2010) kann die Festlegung auf Dreiwortsequenzen einerseits forschungsökonomisch begründet werden. Der gewonnene Datensatz bleibt überschaubar, weil im Verhältnis weniger Dreiwortkombinationen hochfrequent genutzt werden im Vergleich zu Zweiwortkombinationen. Andererseits zeigt sich, dass in Dreiwortkombinationen häufige Zweiwortkombinationen enthalten sind (Simpson-Vlach & Ellis, 2010, S. 491f.). Somit werden auch implizit Informationen über häufige Zweiwortkombinationen geliefert, die bei Entscheidungen zur Vokabularauswahl in UK-Interventionen bei Deutsch als Zweitsprache berücksichtigt werden können. Damit idiosynkratische Äußerungen einzelner Personen erkannt werden können, lässt sich ein Streuungskriterium anwenden (Biber & Barbieri, 2007, S. 268). Allgemeingültige Streuungskriterien konnten in der bisherigen Forschung zu festen Wortkombinationen nicht gefunden werden. Daher wurde das verwendete Kriterium $\geq 50\%$ ($n = 11$) aus der Kernvokabularforschung als *zusätzliches* Beschreibungsmerkmal in Form einer Feinanalyse angewendet.

Obwohl eine quantitative Analysemethode gewählt wurde, kann in Anlehnung an Meißner et al. (2016, S. 309) der Einfluss qualitativer Methoden, im Sinne beschreibender und deutender Herangehensweisen, nicht vollkommen ausgeschlossen werden. »Das qualitative Sichten von Einzelbelegen ist zudem unverzichtbarer Bestandteil zur Validierung von quantitativen Analysen, welche zumindest auf der Formebene ansetzen (müssen) und semantisch begründete Ambiguitäten (Homonymie, Polysemie und kontextspezifische Bedeutungsvarianten) systematisch ausblenden« (Meißner et al. 2016, S. 309). Mit Blick auf das gegenwärtige Forschungsinteresse wird auf eine detaillierte Auswertung von Homonymen und kontextspezifischen Bedeutungsvarianten verzichtet. Auf Kommunikationshilfen steht nur eine begrenzte Anzahl an Feldern zur Verfügung. Um Klickwege zu reduzieren und motorische Automatisierungsprozesse zu unterstützen, sollte das Angebot von Homonymen und kontextspezifischen Bedeutungsvarianten sorgfältig abgewogen werden (auch Sachse et al., 2013).

Erkenntnisinteresse

Ziel der Studie war es, das Kernvokabular sowie feste Wortkombinationen in der frühen Phase der Aneignung von Deutsch als Zweitsprache bei Kindern in der Primarstufe zu untersuchen.

Folgende *Teilziele* waren mit der Untersuchung verbunden:

- Betrachtung des natürlichen mündlichen Sprachgebrauchs in Bezug auf den produktiven Wortschatz (Token, Lemma-Types) pro Untersuchungsteilnehmer:in.

- Bestimmung des Kernvokabulars pro Untersuchungsteilnehmer:in nach der 80 %-Marke (Boenisch, 2014b).
- Erstellung einer Gesamtliste zum Kernvokabular: Bestimmung der 80 %-Marke (Boenisch, 2014b) (*Umfang des Kernvokabulars*); Analyse des Kernvokabulars im Hinblick auf Inhalts- und Funktionswörter sowie Wortartenverteilung (*Zusammensetzung des Kernvokabulars*).
- Erstellung einer Gesamtliste zu den Dreiwortkombinationen: Analyse der festen Wortkombinationen
- Vergleich der Kernvokabularliste mit der Liste über die festen Wortkombinationen

Ahrenholz (2014) beschreibt, dass bei der Analyse von Lernersprachen Vergleichsdaten berücksichtigt werden sollten (S. 173, auch Crossley & Salsbury, 2011, S. 3, Interlanguagehypothese Kap. 2.3). »Orientierungsgröße können immer nur Äußerungen von Muttersprachlern sein, die ungefähr gleich alt sind (gilt besonders im frühen Zweitspracherwerb), gleiche oder vergleichbare kommunikative Aufgaben erhalten haben und diese im gleichen Medium (mündlich/schriftlich) unter vergleichbaren Kontextbedingungen lösen mussten« (Ahrenholz, 2014, S. 175).

Daher wurde ein Teildatensatz (Grundschule Klasse 2 und 4, Boenisch, 2013) aus der Studie von Boenisch (2014b) als *Vergleichsdatensatz* (Referenzkorpus) herangezogen. Ziel war es, Gemeinsamkeiten und Unterschiede im Kernvokabular sowie bei den festen Wortkombinationen zu identifizieren.

Begründung für die Unterstützte Kommunikation

Der Untersuchungskorpus (Primärkorpus) spiegelt den Sprachgebrauch von Kindern wider, die Deutsch als Zweitsprache in der frühen Phase erwerben. Der produktive Wortschatz der Kinder kann zunächst als noch sehr reduziert eingeschätzt werden. Daher können über die Korpusanalyse primär die Wörter und festen Wortkombinationen herausgearbeitet werden, die zu *Beginn des Erwerbs von Deutsch als Zweitsprache* kommunikativ bedeutungsvoll sind. Ein solcher Wortschatz kann dann genauso in der frühen Kommunikationsförderung von Bedeutung sein. Die Versorgungssituation mehrsprachig aufwachsender Kinder mit UK-Bedarf ist in Deutschland noch ausbaufähig (Kap. 3.2). Somit fehlen auch valide Datensätze, um den Sprachgebrauch mehrsprachig unterstützt kommunizierender Personen umfänglich zu erheben. Auf Grundlage des Sprachgebrauchs neuzugewanderter Kinder mit noch geringen Deutschkenntnissen lassen sich erste Tendenzen hinsichtlich der Erwerbsbedingungen ableiten, die auf unterstützt kommunizierende Kinder übertragen werden können. Eine Analyse der Erwerbsbedingungen von unterstützt kommunizierenden Kindern mit Deutsch als Zweitsprache lässt sich aufgrund der fehlenden Lautsprache und des eingeschränkten, vorgegebenen Vokabulars auf den Kommunikationshilfen nur in begrenztem Maße anfertigen. Anhand des Erkenntnisinteresses sollen keinesfalls heterogene Entwicklungsverläufe aberkannt werden oder normative Vorgaben etabliert werden (siehe auch Tönsing & Soto, 2020, S. 198). Vielmehr werden die »Entwicklungsanalogien« in Anlehnung an Haupt (1996, S. 18) von Kindern mit und ohne Beeinträchtigung anerkannt (auch Schäfer, Zentel & Manser, 2022, S. 22ff.). Tiefere Einsichten in den natürlichen

mündlichen Sprachgebrauch sollen gewonnen werden und als ergänzende Ressource im komplexen Prozess der Vokabularauswahl bei mehrsprachigen UK-Interventionen bereitgestellt werden (Tönsing & Soto, 2020, S. 198, Kap. 6).

10.2 Begründung der Forschungsmethode

Jede zugrunde gelegte Theorie, das Forschungsinteresse, Forschungstraditionen und der Kenntnisstand haben Auswirkungen auf methodische Entscheidungsprozesse (Higginbotham, 2003, S. 54; Schramm & Marx, 2017, S. 211). Die vorliegende Arbeit ist in den sprachtheoretischen Diskurs gebrauchsbasierter Ansätze eingebettet (Kap. 4). Das Erkenntnisinteresse bestand vor allem darin, die Performanz, »[d]as, was tatsächlich geäußert und beobachtet werden kann« (Perkuhn et al., 2012, S. 11) im Hinblick auf die am häufigsten genutzten Wörter (Kernvokabular) und festen Wortkombinationen zu untersuchen. Daher wurde die *Korpuslinguistik* als Methodologie herangezogen. Korpusbasierte Analysen erlauben es, Aussagen über die häufigkeitsbasierte Verteilung von Wörtern in einem ausgewählten Korpus sowie Äußerungen mit Fokus auf einzelne Wörter oder die jeweiligen Kombinationsmöglichkeiten zu untersuchen (Perkuhn et al., 2012, S. 78ff., 110ff.; Wray & Perkins, 2000, S. 7). »Korpuslinguistik untersucht *Sprache im Gebrauch*« (Perkuhn et al., 2012, S. 14). Das Erkenntnisinteresse bezog sich auf die »Erfassung« (Mezger, Schellhardt & Şimşek, 2016, S. 182) des Sprachgebrauchs. Daher wurde kein gezielter Stimulus als Sprech Anlass gesetzt, um spezifische sprachliche Äußerungen zu provozieren (Mezger et al., 2016, S. 182). Die Betrachtung von Korpora werden als »das beste Verhältnis von Praktikabilität zu Natürlichkeit der Daten« (Shadrova, 2013, S. 20) erachtet.

Mit der vorliegenden Arbeit war *nicht* das Anliegen verbunden, Rückschlüsse auf den Erwerb bestimmter sprachlicher Regeln oder sprachlicher Fertigkeiten zu ziehen (Engemann, 2018, S. 110; Mezger et al., 2014, S. 73). Deshalb wurde auf die Untersuchung spezifischer kommunikativer Kontexte, wie z.B. Schule, Elternhaus, Freizeitbereich (Ahrenholz, 2014, S. 176) sowie auf die Einbeziehung standardisierter und formeller Sprachtests verzichtet (Überblick bei Porsch, 2014, S. 87ff.). Speziell bei der Erfassung fester Wortkombinationen wurde die Bewertung (ziel)sprachlicher Korrektheit vernachlässigt. In Anlehnung an Aguado (2002) gilt das Merkmal der Korrektheit zur Identifizierung von festen Wortkombinationen als irrelevant (S. 60).