

Swain, Merrill

- 2000 The output hypothesis and beyond: mediating acquisition through collaborative dialogue. In: James P. Lantolf (Hg.), *Sociocultural Theory and Second Language Learning*, 97–114. Oxford: University Press.

Swain, Merrill

- 2005 The Output Hypothesis: Theory and Research. In: Eli Hinkel (Hg.), *Handbook of Research in Second Language Teaching and Learning*, 471–483. Mahwah: Erlbaum.

Tschirner, Erwin

- 1996 Spracherwerb im Unterricht: Der *Natural Approach*. *Fremdsprachen lehren und lernen* 25: 50–69.

Wardhaugh, Ronald

- 1970 The contrastive analysis hypothesis. *TESOL Quarterly* 4: 123–130.

Frank G. Königs, Marburg (Deutschland)

85. Empirische Forschungsmethoden in der Zweit- und Fremdsprachenerwerbsforschung

1. Einleitung
2. Quantitative und qualitative Forschung
3. Triangulation
4. Gütekriterien
5. Methoden der Datenerhebung
6. Methoden der Datenaufbereitung, -auswertung und -interpretation
7. Ausblick
8. Literatur in Auswahl

1. Einleitung

Empirische Zweit- und Fremdsprachenerwerbsforschung sammelt datengeleitet, systematisch und methodisch kontrolliert Erkenntnisse über die Wirklichkeit des Lehrens und Lernens von Zweit- und Fremdsprachen (hier: DaF und DaZ). Dabei greift sie forschungsmethodologische Prinzipien und Forschungsmethoden aus Nachbardisziplinen auf, insbesondere aus Psychologie und Sozialwissenschaften, und ergänzt diese um Verfahren der (angewandten) Sprachwissenschaften. Forschung im Bereich DaF/DaZ kann dabei an Entwicklungen in der internationalen Fremdsprachenforschung (insbesondere *Applied Linguistics*, *Second Language Acquisition Research*, Bilingualismusforschung) und in der deutschen Fremdsprachenforschung (Sprachlehr- und -lernforschung, empirische Fremdsprachendidaktik) anknüpfen. In Deutschland wurde parallel zur Konstituierung des Faches DaF an den Universitäten insbesondere im Rahmen der Sprachlehr- und -lernforschung empirische Forschung vorangetrieben. Ein wichtiges Ziel dabei ist stets, „die begründete Konsolidierung bzw. Veränderung konkreter Formen des Lehrens

und Lernens fremder Sprachen zu bewirken“ (Bausch, Christ und Krumm 2003: 4). Dies impliziert, „die Forschungsmethodik so anzulegen, dass sie Probleme aus der Praxis aufgreift, der systematischen und integrativen Forschung zuführt und wieder in die Praxis einbringt, sei es in Form von Bestätigungen für gewohntes Unterrichtsverhalten, sei es als Empfehlung bzw. Handlungsalternative für eine begründete Veränderung desselben“ (Bausch, Christ und Krumm 2003: 4). Empirische Forschung im Bereich DaF und DaZ ist also auch dem Postulat verpflichtet, dass sie über die allgemeine Erkenntnisgewinnung hinaus zumindest mittelbar für die Praxis relevant ist. Neben der beschriebenen angewandten Forschung haben auch Grundlagen- und Aktionsforschung (vgl. Art. 153) ihren festen Platz in der Empirie des Faches (vgl. McDonough und McDonough 1997: 42–43; Seliger und Shohamy 1989: 19–20).

Empirische Forschung im Bereich DaF und DaZ muss sich – wie die L2-Forschung insgesamt – damit auseinandersetzen, dass das Lernen und Lehren von Fremd- und Zweitsprachen durch eine Vielzahl einander wechselseitig beeinflussender Faktoren bestimmt wird (Faktorenkomplexion), dass Lehr- und Lernprozesse innerhalb eines zu bestimmenden sozialen Milieus und ggfs. innerhalb von Institutionen stattfinden, dass man es sowohl mit Produkten als auch mit Prozessen zu tun hat und dass diese Prozesse des Lehrens und Lernens dynamisch sowie von individuellen Unterschieden zwischen Lernenden und durch Instabilität der Lernprodukte geprägt sind (vgl. Grotjahn 2003: 493). Hierbei zu treffende forschungsmethodologische Entscheidungen führen zur Verortung eines Forschungsprojekts im Rahmen eines stärker quantitativ oder qualitativ orientierten Forschungsansatzes.

2. Quantitative und qualitative Forschung

In der empirischen Forschung werden zwei forschungsmethodologische Paradigmen unterschieden, die disziplinenübergreifend als *qualitative* und *quantitative* Forschungsansätze bezeichnet werden und die die Enden eines bipolaren Kontinuums markieren (Grotjahn 1987: 59–60; Seliger und Shohamy 1989: 114; Larsen-Freeman und Long 1991: 15; Nunan 1992: 3–8; Brown und Rodgers 2002: 15–16; Riemer 2007). *Methodologie* meint hier eine Metatheorie empirischer Methoden (Methodenlehre), im Unterschied zum Terminus *Methode*, der sich auf einzelne wissenschaftliche Verfahrensweisen im Umgang mit Daten bezieht. Während die Erfahrungsrealität in qualitativer Forschung verbalisiert wird, wird sie in der quantitativen Forschung numerisch beschrieben (Brown und Rodgers 2002: 11–12; 15; Bortz und Döring 2006: 296). Darüber hinaus implizieren qualitative und quantitative Forschung jeweils unterschiedliche Voraussetzungen und Ziele, gehen von unterschiedlichen zugrunde gelegten Menschenbildern und Vorstellungen über die Forschungsorganisation aus und arbeiten mit Daten, die nach unterschiedlichen Maximen erhoben, aufbereitet und analysiert werden. Somit bezieht sich die Kategorisierung *qualitativ – quantitativ* entweder auf die Datenerhebung, auf die Daten selbst und/oder auf die Auswertung der Daten (Grotjahn 1987: 59; Nunan 1992: 4). Dies führt innerhalb der jeweiligen Paradigmen zu Forschungsergebnissen mit unterschiedlichem Erkenntnisanspruch.

Die von Grotjahn (1987: 55–60) präferierte Verwendung der Termini „explorativ-interpretativ“ bzw. „analytisch-nomologisch“ als Bezeichnung dieser beiden Forschungs-

sausrichtungen benennt die zentralen Merkmale beider Methodologien. Beide Ansätze wurden lange Zeit als sich gegenseitig ausschließende Paradigmen interpretiert. Mittlerweile ist der einander ergänzende Wert beider Ansätze unbestritten, wobei eine gewinnbringende Verknüpfung der Herangehensweisen allerdings noch zu selten stattfindet.

2.1. Quantitative Forschung

Beschreibung und *Erklärung* menschlichen Verhaltens (hier: Lernen und Lehren von DaF/DaZ) aus der Außenperspektive ist das Ziel quantitativer L2-Forschung, und zwar idealerweise im Rahmen generalisierbarer Gesetzmäßigkeiten. Aufgestellte Hypothesen sollen dabei mit Hilfe empirischer Untersuchungen verifiziert oder falsifiziert werden. Sie müssen also so formuliert sein, dass deutlich wird, wie sie, auch im Rahmen nachfolgender Replikationsstudien, falsifiziert werden könnten (Seliger und Shohamy 1989: 11). Dazu werden stets Hypothesenpaare aus Nullhypothese und Alternativhypothese aufgestellt, wobei die Alternativhypothese einen Effekt annimmt, den die Nullhypothese negiert. Ein signifikantes Ergebnis bedeutet, dass ein in einer Stichprobe ermittelter Befund es unwahrscheinlich erscheinen lässt, dass in der Population die Nullhypothese gelten könnte. Daher wird automatisch die Alternativhypothese angenommen. Ein nicht signifikantes Ergebnis kann hingegen nicht als Hinweis auf das Zutreffen der Nullhypothese gewertet werden (vgl. Bortz und Döring 2006: 24–27).

Der statistische Kennwert p (für engl. *probability*, Wahrscheinlichkeit) gibt an, mit welcher Wahrscheinlichkeit das in der Stichprobe gefundene Ergebnis für die Population verallgemeinerbar ist. Häufig wird dabei unterschieden in $p \geq 0.05$ (nicht signifikant), $p \leq 0.05^*$ (signifikant), $p \leq 0.01^{**}$ (sehr signifikant) und $p \leq 0.001^{***}$ (hoch signifikant). Signifikanz ist jedoch nicht zu verwechseln mit der Bedeutsamkeit statistischer Ergebnisse. Bei sehr großen Stichproben werden auch sehr geringe Zusammenhänge bzw. Unterschiede signifikant.

Im Rahmen quantitativer Forschung ist es unabdingbar, Teilbereiche des Untersuchungsfeldes zu isolieren und zu kontrollieren, was bedeutet, dass externe Einflüsse und weitere auf die interessierenden Untersuchungsgegenstände intervenierende Variablen hinreichend kontrolliert werden müssen. Dies erfordert in der Regel eine Manipulation des Untersuchungsfeldes, sorgfältige Zusammenstellung der Probandengruppen und standardisierte Datenerhebungs- und -auswertungsverfahren sowie größere Stichproben. Erhoben werden sollen harte – d. h. reliable und replizierbare – Daten, die mit Hilfe teststatistisch überprüfter Instrumentarien elizitiert werden. Dies ist jedoch häufig nur auf Kosten der Tiefgründigkeit und Natürlichkeit und damit der externen Validität quantitativer Daten zu erreichen (s. Abschnitt 4).

Experimente markieren den quantitativ ausgerichteten Extrempol der methodologischen Skala. Unter kontrollierten (Labor-)Bedingungen können Aussagen über Ursache-Wirkungs-Beziehungen durch den weitgehenden Ausschluss von Störvariablen angestrebt werden. Das Experiment definiert sich durch eine Zufallszuweisung (*random sampling*) bzw. zumindest anhand relevanter Merkmale kontrollierte Zuweisung der Untersuchungsteilnehmer auf eine Versuchs- und eine Kontrollgruppe. Die interessierende Einflussvariable wird durch eine von dem Forschenden initiierte Manipulation (z. B. eine spezifische Lehrtechnik) operationalisiert, der nur die Versuchsgruppe ausgesetzt wird, nicht aber die Kontrollgruppe. Das klassische Design eines Experimentes stellt die Ab-

folge Vortest (*pretest*) – plus/minus Manipulation (*treatment*) – Folgetest (*posttest*) dar (vgl. Seliger und Shohamy 1989: 135–152). In der L2-Forschung, wie auch allgemein in den Sozialwissenschaften, sind Experimente aufgrund der schwierig herzustellenden Bedingungen (insbesondere: Zufallsverteilung von Untersuchungsteilnehmern z. B. auf Lerngruppen), aus forschungsethischen Beweggründen (z. B. nur der Hälfte der Probandinnen und Probanden eine Fördermaßnahme angedeihen zu lassen) und aufgrund von Zweifeln an der ökologischen Validität der Befunde nur selten zu finden. Etwas häufiger sind *Quasi-Experimente*, die im Unterschied zu Experimenten mit natürlich existierenden Gruppen arbeiten. Typischerweise im Rahmen von quantitativen Designs eingesetzte Methoden sind insbesondere Tests oder Fragebogen.

2.2. Qualitative Forschung

Exploration komplexer Prozesse sowie deren *Verstehen* ist das Hauptziel qualitativer Fremdsprachenforschung. Dies verlangt einen empathischen Nachvollzug aus der Perspektive der UntersuchungsteilnehmerInnen und bedingt ein komplexes elaborativ-prospektives Menschenbild, wobei das Gesamtfeld als Informationslieferant zur Verfügung steht (zur Menschenbilddiskussion in der Fremdsprachenforschung vgl. Grotjahn 2005). Hypothesen und Theorien sollen dabei erst während des Forschungsprozesses durch das interpretative Auffinden wiederkehrender Muster erschlossen werden (qualitative Forschung als theoriegenerierende Forschung), wobei ausgeprägte Vorstrukturierungen des Untersuchungsfeldes (z. B. Standardisierungen in der Datenerhebung oder vorrangige Berücksichtigung von Vorwissen in der Entwicklung kategorialer Systeme) weitgehend vermieden werden sollen, was den Einsatz offener Forschungsmethoden impliziert. Es interessieren nicht allein die *Produkte* menschlichen Verhaltens, sondern auch die *Prozesse*, die zu den Produkten führen. Das Untersuchungsfeld soll so weit wie möglich natürlich belassen sein, dies heißt konkret: Soll z. B. gesteuerter Fremdspracherwerb *verstanden* werden, so sind die Daten aus dem Kontext real stattfindenden Fremdsprachenunterrichts zu gewinnen und nicht in speziell eingerichteten laborähnlichen Handlungsräumen mit je eigenen Konstellationen. Dabei sollen möglichst tiefgründige, reichhaltige Daten (*rich data*) erhoben werden. Dies bedeutet andererseits, dass es aus forschungsökonomischen Gründen selten möglich ist, größere Probandengruppen zu erfassen, und dass damit qualitative Forschungen zumeist in Form von Fallstudien organisiert sind, die auf eine Generalisierung der Befunde verzichten müssen. Noch zu wenig ist in der qualitativen Fremdsprachenforschung die Auswahl von Fällen diskutiert worden. Auch im Rahmen eines qualitativen Forschungsdesigns notwendige Entscheidungen für ein bestimmtes Lehr-/Lernumfeld (Alter, Ausgangssprache, L2-Kompetenz, Unterrichtskontext, Lernziele etc.) inklusive Verzicht auf größere Probandengruppen (und repräsentative Stichproben) dürfen nicht zu Beliebigkeit bzw. Zufälligkeit der Auswahl der UntersuchungsteilnehmerInnen führen. Forschungsansätze, die sich an der *Grounded Theory* (vgl. Glaser und Strauss 1998: 53–83) orientieren, arbeiten mit Verfahren des *theoretischen Sampling*, wobei Untersuchungspersonen bzw. -gruppen sukzessiv festgelegt werden und durch die systematische Berücksichtigung zunächst typischer und ähnlicher, dann abweichender und kritischer Fälle eine maximale Variation angestrebt wird. Die dabei entwickelten Erkenntnisse lenken dann bestenfalls die weitere Fallauswahl.

Methoden der *Feldforschung* sind an den Prinzipien qualitativer Forschung im Idealfall sehr nah orientiert, da Untersuchungsgegenstände in ihrem natürlichen Kontext untersucht werden. Die Untersuchungsteilnehmer verbleiben in ihren alltäglichen Praktiken (z. B. in ihren alltäglichen sprachlichen Handlungsräumen), was artifizielle Forschungsergebnisse verhindern soll. Wiederholte teilnehmende Beobachtung und offene Befragungen (s. u.) im Rahmen von Längsschnittstudien (Longitudinalstudien) über einen längeren Zeitraum sind typischerweise dabei eingesetzte Methoden; Forschungsstrategien werden flexibel gehandhabt. Da der Untersuchungskontext nicht manipuliert werden soll/darf, ist die Kontrolle von Variablen nicht in Gänze möglich (eingeschränkte interne Validität), den Untersuchungsergebnissen wird aber – im Gegensatz zu Forschungsergebnissen quantitativer Studien – eine hohe ökologische und externe Validität zugeschrieben.

3. Triangulation

In dem unmöglichen Streben nach gleichzeitiger Kontrolle und Natürlichkeit liegt folglich der Grundkonflikt zwischen quantitativer und qualitativer Forschung. In den letzten Jahren hat sich (nicht nur) in der Fremdsprachenforschung der Standpunkt durchgesetzt, dass der qualitative und der quantitative Forschungsansatz je spezifische Stärken und Schwächen mit sich bringen und dass stets Erkenntnisinteresse und Untersuchungsgegenstände die Wahl spezifischer Forschungsdesigns begründen sollten. So ist es für die Auswahl eines Forschungsansatzes beispielsweise relevant, ob der Wissensstand zum spezifischen Gegenstand eine begründete Hypothesenprüfung im Rahmen eines in der Regel quantitativen Designs erlaubt oder ob es sinnvoller ist, mögliche Vorannahmen und Hypothesen über das Untersuchungsfeld zunächst im Rahmen einer qualitativen Studie zu gewinnen.

Um Synergieeffekte zwischen den beiden beschriebenen Herangehensweisen zu erzielen, ist die *Triangulation* unterschiedlicher Forschungsmethoden und/oder Methodologien im Untersuchungsplan eine in der empirischen Fremdsprachenforschung immer häufiger eingesetzte Forschungsstrategie. Der aus der Navigation entlehnte Begriff bezieht sich auf die Vermessung eines Objekts von unterschiedlichen Standpunkten aus (Brown und Rodgers 2002: 243; McDonough und McDonough 1997: 71; Aguado und Riemer 2001: 247). Zentrale Funktionen von Triangulation sind die Überwindung von Schwächen und Verzerrungen von Forschungsmethoden durch den Einsatz von mehrmethodischen Forschungsansätzen (Validierungsstrategie) und/oder die Suche nach reichhaltigeren Einsichten in den Forschungsgegenstand: „Das Ergebnis einer Triangulation kann Konvergenz, Komplementarität und Divergenz sein“ (Grotjahn 2006: 260). Zentral ist, dass die triangulierten Daten tatsächlich einen *gemeinsamen* Untersuchungsgegenstand betreffen; der Einsatz unterschiedlicher Datensätze ist per se noch keine Triangulation (vgl. Aguado und Riemer 2001).

Von Triangulation wird in zwei zu unterscheidenden Kontexten gesprochen. Zum einen bedeutet Triangulation die Kombination qualitativer und quantitativer Teilstudien im Rahmen ein und desselben Forschungsprojekts, z. B. durch die Verknüpfung einer Stichprobe mit ergänzenden Fallstudien. Auch die sukzessive Abfolge von qualitativer Pilotstudie und quantitativer Querschnittstudie kann hierunter gefasst werden (vgl. Flick

2004: 67–85). Zum anderen wird Triangulation im Rahmen rein qualitativer Forschungsansätze zur Qualitätsverbesserung durch die Verwendung mehrmethodischer Ansätze eingesetzt (vgl. Aguado und Riemer 2001). Unterschieden werden kann weiter zwischen *Methoden-Triangulation* (unterschiedliche Methoden zur Untersuchung desselben Gegenstands: multiple Operationalisierung), *Daten-Triangulation* (unterschiedliche Datenquellen für das gleiche Phänomen: unterschiedliche Orte, Zeitpunkte, Personen unterschiedlicher Herkunft), *Forschenden-Triangulation* (unterschiedliche Beobachter, Interviewer, Datenauswerter etc.) und *Theorien-Triangulation* (Datenanalyse unter Einbeziehung unterschiedlicher theoretischer Ansätze; vgl. Denzin 1978; McDonough und McDonough 1997: 71; Brown und Rodgers 2002: 243–244; Grotjahn 2006: 259–260).

4. Gütekriterien

Gütekriterien sind „explizit gemachte Beurteilungskriterien, die für einen rationalen Diskurs über die Qualität von Forschung erforderlich sind“ (Aguado und Riemer 2001: 246). Im Rahmen qualitativer Forschung sind andere Gütekriterien anzulegen als in quantitativer Forschung. Als aber beiden Paradigmen übergeordnetes Gütekriterium kann *Transparenz* als Bedingung für Nachvollziehbarkeit von Forschung angesehen werden (vgl. Aguado 2000), da diese eine Einschätzung der Güte von Forschungsarbeiten überhaupt erst möglich macht. Bärenfänger und Stevener (2001: 16–24) haben in diesem Zusammenhang einen offenen Kriterienkatalog zu „Datenerhebungsverfahren und ihre[r] Evaluation“ erstellt, in dem sie fordern, folgende zehn Punkte in der Forschungsliteratur grundsätzlich transparent zu machen: Datenerhebungsverfahren, Probanden, Apparatur, Material, Design und Ablauf, Gegenstände, Operationalisierung, Maße, Gütekriterien sowie Durchführbarkeit. Die Autoren versprechen sich von einer dahingehenden stärkeren Standardisierung eine Ökonomisierung des Forschungsprozesses, insofern als auf Bewährtes zurückgegriffen werden kann, und in der Folge eine erhöhte Qualität der fachinternen Forschung (Bärenfänger und Stevener 2001: 25). Auch das Lernpotential für andere, insbesondere jüngere Forscher, das in der Offenlegung von Überlegungen und Schwierigkeiten steckt, ist ein gewichtiges Argument (Aguado 2000: 119–120).

4.1. Gütekriterien quantitativer Forschung

Zwischen den drei klassischen Gütekriterien quantitativer Forschung, Validität, Reliabilität und Objektivität, besteht ein Inklusionsverhältnis, insofern als es ohne Objektivität keine Reliabilität und ohne Reliabilität keine Validität geben kann (Grotjahn 1987: 62–63).

Das Kriterium der *Objektivität* bezieht sich darauf, dass Datenerhebung, -analyse und -interpretation vom jeweiligen Beobachter unabhängig sein sollten, d. h. dass unterschiedliche Beobachter idealerweise zu demselben Ergebnis kommen sollten (Bortz und Döring 2006: 195). Aguado (2000: 122) hält Objektivität in der empirischen Fremdsprachenforschung jedoch grundsätzlich für eine Illusion; erreicht werden könne bestenfalls intersubjektive Nachvollziehbarkeit. Denn in der Tat ist zumindest die Durchführungsobjektivität nicht zu prüfen, da man z. B. nicht zwei Versuchsleiter mit denselben Pro-

banden denselben Versuch durchführen lassen kann, um die Ergebnisse anschließend zu vergleichen. In Bezug auf Auswertung und Interpretation der Daten sind intersubjektive Vergleiche hingegen möglich. Eine Gefahr für die Durchführungsobjektivität sind z. B. *Pygmalion-* (auch *Polyanna-* bzw. *Rosenthal-*)*Effekte*: Der Versuchsleiter kann, durch seine Erwartungen bedingt, das Verhalten der Probanden beeinflussen, d. h. verstärken oder abschwächen (Brown 1988: 34).

Das Gütekriterium der *Reliabilität* bezieht sich auf die Forderung, dass ein erneutes Messen mit einer vergleichbaren Stichprobe zum gleichen Ergebnis führen sollte. Reliabilität misst also die Zuverlässigkeit der Messmethode (Bortz und Döring 2006: 196). Dabei gilt, wie bereits angesprochen, dass mangelnde Objektivität automatisch auch die Reliabilität senkt, weil durch Unterschiede zwischen Testanwendern Fehlervarianz erzeugt wird (Bortz und Döring 2006: 200).

Validität gilt gemeinhin als das höchste der Gütekriterien. Forschung gilt dann als valide, wenn eine andere Messmethode zu den gleichen Ergebnissen geführt hätte. Validität bezieht sich also auf die Operationalisierung, d. h. die Messbarmachung theoretischer Begriffe in empirisch beobachtbare Phänomene. Anders gesagt geht es um die Frage: Misst das Instrument wirklich, was es messen soll? An dieser Stelle wird deutlich, warum Validität als *conditio sine qua non* gilt. Auch eine objektive und reliable Messung kann invalide sein, wenn sie nicht das misst, was sie messen soll oder zu messen vorgibt (Bell 1999: 104).

Zu unterscheiden sind u. a. *interne* und *externe* Validität. Intern valide sind Ergebnisse dann, wenn sie eindeutig zu interpretieren sind. Die externe Validität bezieht sich auf die Frage, inwiefern die Ergebnisse über die Stichprobe hinaus zu verallgemeinern sind (Seligman und Shohamy 1989: 105; Nunan 1992: 15; Bärenfänger und Stevener 2001: 23; Bortz 2005: 8). Interne und externe Validität stehen dabei grundsätzlich in einem Wechselverhältnis zueinander. Je stärker die Erhebungssituation kontrolliert wird, desto mehr steigt die interne Validität der Messungen. Gleichzeitig sinkt jedoch die Verallgemeinerbarkeit der Ergebnisse, also die externe Validität, da die Stichprobe in der künstlichen Erhebungssituation immer weniger mit der natürlichen Gesamtpopulation gemeinsam hat (Brown 1988: 40; Nunan 1992: 15; Bärenfänger und Stevener 2001: 17, 21; vgl. auch Abschnitt 4.2).

Ferner kann unterschieden werden in *Inhaltsvalidität/Kontentvalidität* (angemessene Operationalisierung des Konstrukts, z. B. kommunikative Kompetenz), *kriterienbezogene Validität* (Korrelation mit einem unabhängigen Außenkriterium, z. B. einem anderen Test), *Augenscheinvalidität* (*face validity*; Akzeptanz des Verfahrens von Seiten der Getesteten und der Testenden) und *Konstruktvalidität* (Frage nach möglichst vollständigen Erfassung des Konstrukts) (vgl. Grotjahn 2000: 312).

4.2. Gütekriterien qualitativer Forschung

Die Orientierung an Gütekriterien wie Offenheit im Forschungsprozess, Einbezug der UntersuchungsteilnehmerInnen, ausreichende Präzisierung des Untersuchungsgegenstandes, Gegenstandsangemessenheit bei der Methodenauswahl, Nachvollziehbarkeit der Datenerhebung, -aufbereitung und -interpretation und Intersubjektivität bzw. reflektierte Subjektivität der Forschungsergebnisse sind Kennzeichen qualitativer Forschung. Im Rahmen der qualitativen Sozialwissenschaften liegt diesbezüglich eine differenzierte Dis-

kussion von Steinke (1999) vor. Hier werden u. a. die vollständige Offenlegung und Dokumentation der (Vor-)Annahmen, Daten und Interpretationen sowie der methodischen Vorgehensweisen (Kriterium der *Indikation*) und im Forschungsprozess auftretender Schwierigkeiten und Widersprüchlichkeiten als Qualitätsnachweis verlangt, des Weiteren die Überprüfbarkeit der empirischen Verankerung der Theoriegenese durch ausreichende Datenbeispiele. Vergleichbare Überlegungen zu Gütekriterien qualitativer Forschung liegen für die Fremdsprachenforschung bzw. DaF/DaZ durch die Arbeitsgruppe Fremdspracherwerb Bielefeld (1996), Aguado (2000) und Richards (2009) vor. Weitreichende Transparenz der Forschungsdokumentation wird auch als Voraussetzung für die Ergänzung durch und Vergleichbarkeit mit komplementären qualitativen Studien gesehen, die der Übertragbarkeit von auf nur kleine Stichproben bezogenen Resultaten auf weitere Kontexte dienlich sein können. Auf diese Weise kann durch gemeinsame Forschungsanstrengungen innerhalb der qualitativen Fremdsprachenforschung zumindest die Bildung von Theorien mittlerer Reichweite angestrebt werden.

5. Methoden der Datenerhebung

Empirische Forschung im Fach DaF/DaZ verwendet vielfältige Methoden zur Datenerhebung, darunter Verfahren der schriftlichen und mündlichen Befragung, Beobachtung, Introspektion und Tests. Im Rahmen quantitativer Forschung sind sowohl die Verfahren der Datenerhebung als auch die Verfahren der Datenaufbereitung und Datenanalyse einer strikten Standardisierung zu unterziehen, während bei qualitativer Forschung offene und semi-offene Verfahren der Datenerhebung und interpretative Verfahren die Datenanalyse bestimmen. Quantitative Daten werden in der Regel im Rahmen stärker elizitierender Kontexte gewonnen, während qualitative Daten im Idealfall nicht oder nur wenig elizitierenden Verfahren entstammen. Damit ist der Datentyp ein Hauptindikator für ein qualitatives oder quantitatives Forschungsdesign. Allerdings sind qualitative Daten nur idealtypisch natürlich und nicht eliziert, denn auch offene Datenerhebungsmethoden strukturieren das Untersuchungsfeld neu. (Wo würde sich beispielsweise ein Lerner des Deutschen als Zweit- oder Fremdsprache in seiner natürlichen Umgebung, dazu vielleicht noch in der Fremdsprache, explizit zu mentalen Prozessen der Informationsverarbeitung äußern außer in einem introspektiven Interview?) Es ist in diesem Zusammenhang anzumerken, dass Daten, die auf Auskünften der UntersuchungsteilnehmerInnen über sie betreffende Belange beruhen (die sogenannten *self report*-Daten), in ihrem Wahrheitsgehalt generell problematisch sind, seien sie nun mit Hilfe offener oder standardisierter Methoden erhoben.

5.1. Beobachtung

Jede Methode der Datenerhebung beruht auf Beobachtung bzw. Sinneswahrnehmung im weiteren Sinne. *Beobachtung* im engeren Sinne ist im Gegensatz zur Alltagsbeobachtung zielgerichtet, systematisch und methodisch kontrolliert, wodurch eine präzisere Wahrnehmung ermöglicht wird. Durch die Festlegung genau zu befolgender Beobachtungsre-

geln kann die wissenschaftliche Beobachtung darüber hinaus intersubjektiv nachvollzogen werden (vgl. Bortz und Döring 2006: 262–263).

Beobachtung ist eine insbesondere im Rahmen von Unterrichtsforschung (*classroom research*) häufig eingesetzte Methode. Beobachtungen können grundsätzlich frei, halbstandardisiert oder standardisiert erfolgen. Damit verbunden kann eher global oder auch sehr selektiv beobachtet werden. Weiter wird zwischen teilnehmender vs. nicht-teilnehmender sowie offener vs. verdeckter Beobachtung unterschieden (vgl. Mayring 2002: 80; Bortz und Döring 2006: 267). Schließlich kann Unterrichtsbeobachtung synchron oder auf Grundlage eines Mittschnitts zeitversetzt und mit der Möglichkeit wiederholter Betrachtung stattfinden. Außer in gesteuerten Unterrichtskontexten findet Beobachtung auch in ungesteuerten Settings Anwendung, so z. B. zur Sprachstandsdiagnostik im Vorschulalter (z. B. SISMIK, vgl. Art. 146).

Bortz und Döring (2006: 265–266) nennen Selektion, Abstraktion, Klassifikation, Systematisierung und Relativierung als die wesentlichen methodischen Probleme der Beobachtung. Das, was vom beobachteten Unterricht letztlich ausgewertet und interpretiert wird, stellt eine starke Reduktion der eigentlichen Aktion statt (vgl. Brown und Rodgers 2002: 101). Während der Datenerhebung sollte daher möglichst breit beobachtet werden und der engere Fokus sollte erst im Rahmen der Datenanalyse angelegt werden (Mackey und Gass 2005: 211). Krumm (2001: 1145–1146) schlägt außerdem vor, dem Problem der (notwendigen) Selektivität von Beobachtungen durch Inter-Rater, Beobachtertrainings, Aufzeichnung der Beobachtung zum Zwecke der wiederholten Betrachtung, Einbezug anderer am Prozess Beteiligter (z. B. der Lehrkräfte) und wiederholte Beobachtung anstelle von Momentaufnahmen zu begegnen.

Weitere methodische Probleme, insbesondere von Unterrichtsbeobachtung, sind darauf zurückzuführen, dass nach wie vor unklar ist, was guten Unterricht ausmacht, dass die vorhandenen Studien häufig nicht miteinander vergleichbar sind (*stand-alone studies*), da die Unterrichtskontexte stark variieren, und dass Unterrichtsbeobachtung relativ zeitintensiv (Planung, Vorbereitung, Durchführung, Transkription, Codierung, Interpretation) ist (Brown und Rodgers 2002: 98–99). Außerdem treten in den Daten häufig Beobachtereffekte auf: Die Anwesenheit eines Beobachters kann die Ergebnisse beeinflussen (Beobachterparadox), was insbesondere problematisch ist, wenn sich dadurch die Qualität des Unterrichts verschlechtert. Einen möglichen Ausweg bieten Längsschnittstudien, die wiederholte Beobachtung erlauben (Mackey und Gass 2005: 187–188; vgl. auch Bortz und Döring 2006: 266).

5.2. Introspektion

Mentale, also nicht beobachtbare Prozesse beim Lehren und Lernen von Fremd- und Zweitsprachen sind kaum zugängliche, nur schwer operationalisierbare Untersuchungsgegenstände, wobei introspektive Verfahren den Versuch einer Annäherung darstellen. Introspektion elizitiert Verbalisierungen von den Untersuchungsteilnehmern über ihre selbst beobachteten Gedankengänge, was mittels unterschiedlicher Instrumente unterstützt wird. Auch im Rahmen von Befragungen (s. Abschnitt 5.3) und Tagebüchern können introspektive Daten erhoben werden, z. B. im Rahmen von lernbiographischen Erzählungen. Von *Introspektion* im engeren Sinne wird allerdings dann gesprochen, wenn Verbalprotokolle während oder (möglichst zeitnah) nach der interessierenden mentalen

Aktivität (z. B. beim Lösen fremdsprachlicher Aufgaben, im Rahmen von Korrektursequenzen) erhoben werden. Prominent eingesetztes Instrument ist das *Laute Denken* bzw. die Erhebung sogenannter *Lautes-Denken-Protokolle*, die simultan zur Ausführung der fokussierten Handlungen und Problemlöseprozesse erfolgen. Die Untersuchungsperson wird dabei gebeten, ihre Gedanken frei und ungefiltert wiederzugeben; Untersuchungsleitern kommt lediglich die Funktion der Initiierung und Aufrechterhaltung des Lauten Denkens zu. *Retrospektionen* dagegen erfolgen zeitlich versetzt: Untersuchungsteilnehmer werden gebeten, ihre Gedanken während des Ausübens einer bestimmten Handlung zu erinnern und zu verbalisieren, auch mit medialer Unterstützung (z. B. Vorspielen der einschlägigen Unterrichtssequenz), was auch als *stimulated recall* bezeichnet wird. Introspektive Datenerhebungen sind (nicht nur) in der Fremdsprachenforschung umstritten, da mit ihnen nur unvollständig kognitive Prozesse abgebildet werden können und mit ihnen die Gefahr einhergeht, die Untersuchungsteilnehmer zu überfordern oder zu Gefälligkeitsaussagen zu ermuntern – bis hin zum Vorwurf an Lautes-Denken-Protokolle, sie würden die eigentlich interessierenden mentalen Prozesse so verändern, dass keine reliable und valide Einsicht möglich sei (vgl. Heine 2005). Introspektive Daten sind immer qualitative Daten, die im weiteren Forschungsverlauf im Rahmen interpretativer Datenanalyse ausgewertet werden.

5.3. Befragung

Es kann zwischen zwei Arten der Befragung unterschieden werden: der schriftlichen und der mündlichen. Schriftliche Befragungen mittels standardisierter Fragebogen werden häufig zur Hypothesentestung eingesetzt und enthalten dementsprechend vorwiegend geschlossene Fragen, d. h. Fragen mit vorgegebenen Antwortkategorien. Fragebogen bieten im Vergleich zu mündlichen Befragungen den Vorteil ehrlicherer, weil anonymere Antworten, die darüber hinaus auch durchdachter sein können, möglicherweise aber gerade aufgrund der anonymen Erhebungssituation auch nur flüchtig und wenig tiefgründig von den Untersuchungsteilnehmern bearbeitet werden. Standardisierte Datenerhebung und -auswertung sind weniger aufwändig als bei mündlichen Befragungen, v. a. im Rahmen qualitativer Interviews; dafür ist die Fragebogenentwicklung sehr zeitaufwändig. Interviews enthalten prototypisch eher offene Fragen, die der Hypothesengenerierung dienen sollen und die qualitativ analysiert werden, auch wenn die Form von Interviews grundsätzlich wiederum von offen/narrativ über semistrukturiert bis zu geschlossen reichen kann. Die Stärken des Verfahrens liegen darin, dass Mündlichkeit für viele (insbesondere nichtmuttersprachliche) Befragte einfacher als Schriftlichkeit ist, klärende Nachfragen beiderseits möglich und die Befragten aufgrund der kommunikativen Gesprächssituation in der Regel motivierter sind.

Eine Variante des Interviews sind das Gruppeninterview bzw. die Gruppendiskussion, an der mehrere Personen als Befragte bzw. Diskutanden beteiligt sind, die sich in ihren Meinungen zu jeweils die Gruppe betreffenden Themen kontrovers oder komplementär verhalten.

Weiter sind im Zusammenhang mit geschlossenen vs. offenen Fragen einige methodische Vor- und Nachteile zu nennen: Geschlossene Fragen bergen einige praktische Vorteile, wie z. B. ihre schnelle Auswertbarkeit oder auch die geringeren Anforderungen an das Ausdrucksvermögen der Befragten. Vor allem aber liegen ihre Stärken in der Ver-

gleichbarkeit der Antworten und der geringeren Wahrscheinlichkeit des Auftretens von Befrager-Fehlern. Offene Fragen hingegen elizitieren tiefgründigere Antworten, da sie dem Befragten ermöglichen, in seinem eigenen Bezugssystem zu antworten, was zum einen Verzerrungen in den Antworten, Nichtantworten und Falschangaben minimiert, zum anderen keine Antwortkategorie von vornherein ausschließt (vgl. McDonough und McDonough 1997; Dörnyei 2003; Bortz und Döring 2006; Porst 1996, 2008). Halboffene Fragen, d. h. standardisierte Fragen mit vorgegebenen Antwortkategorien plus zusätzlicher offener Antwortmöglichkeit, wie z. B. „Sonstiges“, bieten „[...] sich immer dann an, wenn das tatsächliche Universum möglicher Antworten auf eine Frage zwar gut abgeschätzt (geschlossene Frage), aber nicht definitiv bestimmt werden kann (offene Frage)“ (Porst 1996: 739).

Fragebogen können zur Erhebung von personenbezogenen Daten (z. B. Sprachlernbiographie) und/oder zur Erhebung der Kernvariablen (z. B. Einstellungen zum Zielland) eingesetzt werden. Fragebogen-Items können als Frage oder als Statement formuliert werden und werden im Falle von geschlossenen Fragen meist durch Ankreuzen eines Skalenpunktes beantwortet. Die Skalen können verbalisiert oder endpunktbenannt sein, wobei die Endpunktbenennung von Vorteil ist, da sie eine Gleichabständigkeit der Skalenpunkte sicherstellt. Je nach Fragestellung ist auch zu entscheiden, ob eine gerade oder eine ungerade Skala verwendet werden soll, d. h., ob die Antwortskala eine Mittelkategorie aufweisen sollte oder nicht. Zusätzlich kann es notwendig sein, die Möglichkeit zu geben, mit „weiß nicht“ oder Ähnlichem zu antworten, um die Antworten nicht unnötig zu verfälschen. Praktische Tipps zu Aufbau und Gestaltung von Fragebogen geben z. B. Porst (1996, 2008) und Dörnyei (2003). Eine differenzierte Einführung in die Durchführung (qualitativer) Interviews bietet Helfferich (2004). Zentral für beide Arten der Befragung ist vor allem, dass die Fragen verständlich sein und die Befragten nicht in ihrem Antwortverhalten beeinflussen sollen, was prinzipiell durch eine Pilotierung der Befragungsinstrumente zu überprüfen ist. Ausführlichere Hinweise zur Frageformulierung finden sich beispielsweise bei Bortz und Döring (2006: 244–245), McDonough und McDonough (1997: 177), Dörnyei (2003: 53–55) und Porst (2008: 95–114).

Typischerweise auftretende methodische Probleme der Befragung sind Positionseffekte (z. B. erhöhte Zustimmung zu den erstgenannten Elementen einer schriftlichen Liste), Selbstdarstellungseffekte (z. B. Nichtzugeben sozial unerwünschten Verhaltens), Übergeneralisierungen (Halo-Effekte; z. B. undifferenziert positive Bewertung einer beliebten Lehrkraft), *Non-Opinions* (nicht jeder Befragte hat zu allem auch tatsächlich eine Meinung) und Akquieszenz (allgemeine Tendenz, eher zuzustimmen als abzulehnen; Brosius, Koschel und Haas 2008: 99–103; Dörnyei 2003: 13; Schnell, Hill und Esser 2005: 354–355).

5.4. Test

Eine sehr häufig im Rahmen quantitativer Designs eingesetzte Datenerhebungsmethode ist der Test: „Ein Test ist ein wissenschaftliches Routineverfahren zur Untersuchung eines oder mehrerer empirisch abgrenzbarer Persönlichkeitsmerkmale mit dem Ziel einer möglichst quantitativen Aussage über den relativen Grad der individuellen Merkmalsausprägung“ (Lienert und Raatz 1998: 1). Tests im Bereich DaF/DaZ können der Sprachstandsmessung (vgl. auch Art. 143) oder der Messung eines anderen Verhaltens, aber auch der

kontrollierten Elizitierung einzelner Phänomene dienen. Zu unterscheiden sind kompetenz- vs. performanzorientierte Tests sowie norm- vs. kriteriumorientierte. Kompetenzorientierte Tests werden für Inferenzschlüsse auf ein Kriterium (z. B. Gespräch mit einem Patienten in der L2 Deutsch) genutzt (McNamara 2000: 7–8). Normorientierte Tests, die z. B. der Sprachstandsbestimmung dienen, sollten zusätzlich zu den genannten drei zentralen Gütekriterien quantitativer Forschung dem der Normierung entsprechen. Durch die Standardisierung eines Tests und die anschließende Festlegung von Normwerten auf Basis einer Eichstichprobe wird es möglich, individuelle Testwerte ins Verhältnis zur Zielpopulation zu setzen (vgl. Grotjahn 2000: 317).

Auch Tests können – wie Formen der Befragung – wiederum offene (z. B. einen Brief schreiben, SOPI [*stimulated oral proficiency interview*, vgl. Grotjahn 2000: 336–337]), halboffene (z. B. C-Test, *discourse completion task*) und geschlossene (z. B. *Multiple-Choice-Test*, *elicited imitation*) Aufgaben beinhalten (Albers und Bolton 1995: 27–30; vgl. auch Chaudron 2003 für einen Überblick).

Darüber hinaus werden Tests in der empirischen Fremdsprachenforschung auch zur Messung von personalen Variablen, z. B. von Fremdsprachenlerneignung, Lernstilen und weiteren Persönlichkeitsvariablen eingesetzt, wobei häufig in der Psychologie erprobte Verfahren adaptiert werden (vgl. Art. 94).

6. Methoden der Datenaufbereitung, -auswertung und -interpretation

Bevor sie analysiert und interpretiert werden können, müssen Daten in eine bearbeitbare Form gebracht werden. Quantitative Daten werden in der Regel in numerische Werte überführt (quantifiziert) und dann statistisch ausgewertet. Dazu kann im Fall von Sprachdaten im Vorfeld eine Annotation oder Transkription nötig sein. Daten aus qualitativen mündlichen Befragungen und Introspektionen sowie Audio-Mitschnitten und Videographien (z. B. Unterrichtseinheiten, Gespräche in der Zweit-/Fremdsprache) werden verschriftet (transkribiert), wobei je in Abhängigkeit von den Untersuchungsgegenständen eher grobe oder feine Transkriptionen erforderlich werden, die die Charakteristika mündlicher Sprache (inkl. Abbrüche, Selbstkorrekturen, Hörsignale, Pausen) angemessen wiedergeben (vgl. Dittmar 2004). Hilfreiche Dienste beim Transkribieren leisten zunehmend EDV-gestützte Verfahren, z. B. die im Rahmen des Hamburger Sonderforschungsbereichs „Mehrsprachigkeit“ entwickelte kostenfreie und über das Internet herunterladbare Software EXMARALDA, die einen Partitureditor und unterschiedliche ausdifferenzierte Werkzeuge zur sprachwissenschaftlichen Datenanalyse bereithält (vgl. <http://www.exmaralda.org/>).

6.1. Quantitative Datenanalyse

Quantitative Datenanalyse erfolgt mit Hilfe statistischer Verfahrensweisen und zumeist softwareunterstützt, z. B. mittels Excel oder SPSS. Kennwerte der deskriptiven Statistik, d. h. Maße der zentralen Tendenz, wie z. B. Mittelwert (*mean*), und Dispersionsmaße, wie z. B. Standardabweichung (*standard deviation*) und Spannweite (*range*), vermitteln einen

Überblick über die Stichprobe(n). Inferenzstatistische Verfahren testen, inwieweit eine Verallgemeinerung von in der Stichprobe aufgetretenen Phänomenen auf die Grundgesamtheit einer Population zulässig ist.

In Abhängigkeit von der Skalierung der Daten (nominal, ordinal oder metrisch; vgl. z. B. Rasch, Frieze, Hofmann und Naumann 2006: 1–28), der Anzahl von Variablen und nicht zuletzt der Forschungsfrage werden unterschiedliche Rechenverfahren angewendet. Zu unterscheiden ist zwischen uni-, bi- und multivariaten Verfahren. Diese können wiederum insbesondere Unterschieds- und Zusammenhangshypothesen testen. Für die Fremdsprachenforschung und den DaF-/DaZ-Kontext mit Gewinn lesbare Einführungen in die Statistik bieten z. B. Brown (1988) und Rasch, Frieze, Hofmann und Naumann (2006); multivariate Verfahren behandeln zusätzlich Hatch und Lazaraton (1991).

6.2. Qualitative Datenanalyse

Qualitative Datenanalyse ist eine „Entdeckungsreise“ (Dirks 2007); der analytische Zugriff auf qualitative Daten erfolgt über Interpretationsprozesse. Hierbei wird aus den transkribierten Daten (immer mit regelmäßigem Rückblick auf die Audio- und/oder Videomitschnitte) in einem mehrschrittigen Verfahren der Sinn extrahiert, der hinter den Äußerungen der UntersuchungsteilnehmerInnen steht bzw. stehen könnte. Wiederholte und zeilenweise Lektüre, Nachdenken, das Schreiben von Memos, das Finden und Zusammentragen von zugrundeliegenden Kodes durch Induktion und Deduktion bei der Suche nach wiederkehrenden und bedeutungsvollen Mustern sind typische Handlungen der Interpretierenden (vgl. Silverman 1993). Im Rahmen der *Grounded Theory* (nach Strauss und Corbin 1996) werden Prozesse des offenen Kodierens (kleinschrittiges Kodieren direkt und eng formuliert an den Textstellen), axialen Kodierens (Kategorienbildung durch die Suche nach Zusammenhängen zwischen den Kodes, Finden von Kernkategorien) und selektiven Kodierens (Überprüfen der Kernkategorien) unterschieden, wobei die Interpretierenden mit *theoretischer Sensibilität* an die Daten herangehen sollen. Ein systematisches, stark durch Kodierregeln kontrolliertes zweistufiges Verfahren der induktiven und deduktiven Kategorisierung wird im Rahmen der *qualitativen Inhaltsanalyse* (Mayring 2008) v. a. für die Auswertung größerer Textkorpora vorgeschlagen.

Interpretationen streben den Nachvollzug des Untersuchungsbereichs aus der Perspektive der daran Beteiligten an, sie sind generell gegenüber Einflüssen des Forschenden und des Untersuchungskontextes sensibel. Ein Grundproblem der qualitativen Methodologie ist deshalb, sicherzustellen, dass nicht Vorurteile, Vorwissen etc. die qualitative Datenanalyse so steuern, dass lediglich Muster oder Kategorien induziert werden, die im Kern der Erwartung oder Erfahrung des Forschers entsprechen. Interpretation setzt die gleichzeitige Fähigkeit zur Intuition, Empathie und Selbstkritik (Bereitschaft zur Revision von Interpretationen) voraus, stellt jedoch keine exakten Instrumentarien zur Verfügung, die Objektivität gewährleisten können. Die kontrovers diskutierte Frage, mit wie viel theoretischem Vorwissen Forscher an den Interpretationsprozess herangehen sollen, beantwortet Kelle (1996: 32) wie folgt: „Jeder Versuch, theoretische Konzepte allein aus den Daten emergieren zu lassen, wird letztendlich nur dazu führen, dass der Untersucher im Datenmaterial geradezu ertrinkt.“ Fragen der Dateninterpretation sind für die qualitative Forschungsmethodologie virulent, da sie entscheidend mit Fragen nach der Gültigkeit von Forschung verbunden sind. Softwareunterstützte qualitative Datenanalyse (z. B.

mit ATLAS.ti, MAXQDA), die insbesondere der Datendokumentation und Systematisierung von Kodierprozessen förderlich ist, wird seit einigen Jahren verstärkt durchgeführt.

6.3. Sprachwissenschaftliche Analysen

Zur Operationalisierung sprachbezogener Variablen (die in empirischen L2-Erwerbsstudien häufig die abhängige Variable darstellen) haben sprachwissenschaftliche Analyseverfahren einen prominenten Platz im datenanalytischen Spektrum sowohl quantitativer als auch qualitativer Forschung, wobei ganz unterschiedliche Verfahrensgruppen zum Einsatz kommen (vgl. Ellis und Barkhuizen 2005). *Linguistische Fehleranalyse* untersucht Typen und Ursachen von Abweichungen von der zielsprachlichen Norm. *Interlanguage-Analysen* erweitern den Fokus über den defizitären Ansatz der Fehleranalyse hinaus auf allgemeine Merkmale von Lerner Sprachen, was zum Beispiel Lernerstrategien mit einschließt (vgl. Larsen-Freeman und Long 1991: 56–69). Mittels *Profilanalyse* (z. B. Griebhaber 2006) werden L2-Sprecher anhand bestimmter linguistischer Sprachstandsindikatoren, die innerhalb nicht hintergehbaren Erwerbssequenzen erworben werden müssen, auf unterschiedlichen Spracherwerbsstufen angesiedelt. Häufig arbeiten die genannten Methoden zusätzlich mit *kontrastiven Sprachanalysen* (vgl. Art. 52), um die Sprachdaten mehrsprachiger Probanden besser auswerten zu können.

Neben auf spezifische Fragestellungen bezogenen Analyse kategorien spielen linguistische Maße, denen eine Korrelation mit dem allgemeinen Sprachstand zugeschrieben wird, eine besondere Rolle. So werden im kindlichen Spracherwerb beispielsweise die durchschnittliche Äußerungslänge (*mean length of utterance*, MLU) als Maß der morpho-syntaktischen Komplexität oder der Quotient aus *types* und *token* (*Type-token-Relation*, TTR) als Maß der lexikalischen Diversifizierung berechnet; die Zeit, die bei Plosiven zwischen Verschlusslösung und Einsetzen der Stimmbandschwingung verstreicht (*voice onset time*, VOT), gilt als Indikator für den Grad der Bilingualität von Sprechern.

Diskursanalyse als fächerübergreifend eingesetzte Methode untersucht darüber hinaus sprachliche Interaktionen zwischen L1- und L2-Sprechern oder zwischen L2-Sprechern, wobei beispielsweise Interaktionsmuster, soziale Rollen oder Reparatursequenzen im Fokus des Interesses stehen können (vgl. exemplarisch Henrici 1995).

7. Ausblick

Empirische Fremdsprachenforschung im Bereich DaF/DaZ hat trotz aller Fortschritte noch mit einer Reihe von Unwägbarkeiten zu kämpfen: Längsschnittstudien sind gegenüber Querschnittstudien unterrepräsentiert, was zum Teil auf die schwierige Finanzierung von Forschungsprojekten und auf eingeschränkte zeitliche Ressourcen im Rahmen von Qualifikationsprojekten von NachwuchswissenschaftlerInnen zurückzuführen ist. Einzel Forschungen sind darüber hinaus häufig durch mangelnde Vergleichbarkeit und damit Komplementarität – sowohl auf der Ebene der Methodologie als auch auf der Ebene der Untersuchungsgegenstände – gekennzeichnet, die durch das Fehlen eines allgemein akzeptierten gemeinsamen Bezugsrahmens, aber auch durch die Komplexität des For-

schungsgegenstands Spracherwerb verursacht sein mögen (so ist eine Übertragung von DaF-Studien auf den DaZ-Kontext beispielsweise häufig nicht möglich; vgl. Riemer 2007: 454–455). Replikationsstudien haben bedauerlicherweise Raritätenstatus.

Insgesamt ist eine breiter angelegte forschungsmethodologische Diskussion im Fach ein wichtiges Desiderat. Auch die Ausbildung zum Forscher und zur Forscherin mittels der gezielten Vermittlung von forschungsmethodologischen Kenntnissen und forschungspraktischen Fertigkeiten sollte im Rahmen der DaF- und Germanistik-Studiengänge an den Universitäten innerhalb und außerhalb deutschsprachiger Ländern deutlich verbessert bzw. überhaupt gewährleistet werden. Die Kenntnis grundlegender forschungsmethodologischer Zusammenhänge ist für Studierende jedoch zum einen unabdingbar für die kritische Rezeption empirischer Studien und die Einschätzung der Relevanz von Forschungsergebnissen. Darüber hinaus kann diese Qualifikation auch im Zusammenhang lebenslangen Lernens und für die Arbeit eines reflektierten Praktikers von Nutzen sein. Die vergleichsweise geringe methodologische Qualifikation hat auch zur Folge, dass DaFler/DaZler in aktuellen Bildungsstudien noch deutlich unterrepräsentiert sind, obwohl die Perspektive der Spracherwerbsforschung hier durchaus einen wichtigen Beitrag zum interdisziplinären Spektrum leisten könnte und sollte. Dass im Zuge des Bologna-Prozesses im Rahmen von Master- und Promotionsprogrammen hier breite Verbesserungen erreicht werden können, ist zu hoffen, ist aber auch davon abhängig, dass es gelingt, die Kapazitäten in Forschung und Lehre zu halten und zu erweitern – und dass Forschungsmethodenkompetenz im Bereich der Empirie als zentrales Ausbildungsziel anerkannt wird.

8. Literatur in Auswahl

Aguado, Karin

- 2000 Empirische Fremdsprachenerwerbsforschung. Ein Plädoyer für mehr Transparenz. In: Karin Aguado (Hg.), *Zur Methodologie in der empirischen Fremdsprachenforschung*, 119–131. Baltmannsweiler: Schneider.

Aguado, Karin und Claudia Riemer

- 2001 Triangulation. Chancen und Grenzen mehrmethodischer empirischer Forschung. In: Karin Aguado und Claudia Riemer (Hg.), *Wege und Ziele. Zur Theorie, Empirie und Praxis des Deutschen als Fremdsprache (und anderer Fremdsprachen)*, 245–257. Baltmannsweiler: Schneider.

Albers, Hans-Georg und Sibylle Bolton

- 1995 *Testen und Prüfen in der Grundstufe. Einstufungstests und Sprachstandsprüfungen*. Berlin etc: Langenscheidt.

Arbeitsgruppe Fremdsprachenerwerb Bielefeld

- 1996 Fremdsprachenerwerbsspezifische Forschung. Aber wie? Theoretische und methodologische Überlegungen. Teil I in *Deutsch als Fremdsprache* 33: 144–155; Teil II in *Deutsch als Fremdsprache* 33: 200–210.

Bärenfänger, Olaf und Jan Stevener

- 2001 Datenerhebungsverfahren und ihre Evaluation. Ein Kriterienkatalog. *Fremdsprachen Lehren und Lehren* 30: 13–27.

Bausch, Karl-Richard, Herbert Christ und Hans-Jürgen Krumm

- 2003 Fremdsprachendidaktik und Sprachlehrforschung. In: Karl-Richard Bausch, Herbert Christ und Hans-Jürgen Krumm (Hg.), *Handbuch Fremdsprachenunterricht*, 1–9. 4. vollst. neu bearb. Aufl. Tübingen/Basel: Francke.

- Bell, Judith
1999 *Doing Your Research Project. A Guide for First-Time Researchers in Education and Social Science*. 3. Aufl. Berkshire: Open University Press.
- Bortz, Jürgen
2005 *Statistik für Human- und Sozialwissenschaftler*. 6., vollst. überarb. und aktual. Aufl. Heidelberg: Springer.
- Bortz, Jürgen und Nicola Döring
2006 *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 4., überarb. Aufl. Heidelberg: Springer.
- Brosius, Hans-Bernd, Friederike Koschel und Alexander Haas
2008 *Methoden der empirischen Kommunikationsforschung. Eine Einführung*. 4., überarb. und erw. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Brown, James D.
1988 *Understanding Research in Second Language Learning. A Teacher's Guide to Statistics and Research Design*. Cambridge etc.: Cambridge University Press.
- Brown, James Dean und Theodore S. Rodgers
2002 *Doing Second Language Research*. Oxford: Oxford University Press.
- Chaudron, Craig
2003 Data collection in SLA research. In: Catherine J. Doughty und Michael H. Long (Hg.), *The Handbook of Second Language Acquisition*, 762–828. Malden/MA: Blackwell.
- Denzin, Norman K.
1978 *The Research Act. A Theoretical Introduction to Sociological Methods*. New York: McGraw-Hill.
- Dirks, Una
2007 Fremdsprachenforschung als ‚Entdeckungsreise‘. Im Spannungsfeld von Abduktion, Deduktion und Induktion. In: Helmut Johannes Vollmer (Hg.), *Synergieeffekte in der Fremdsprachenforschung. Empirische Zugänge, Probleme, Ergebnisse*, 43–58. Frankfurt a. M. etc: Lang.
- Dittmar, Norbert
2004 *Transkription. Ein Leitfaden mit Aufgaben für Studenten, Forscher und Laien*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Dörnyei, Zoltán
2003 *Questionnaires in Second Language Research. Construction, Administration, and Processing*. Mahwah/NJ: Erlbaum.
- Ellis, Rod und Gary Barkhuizen
2005 *Analysing Learner Language*. Oxford: Oxford University Press.
- Flick, Uwe
2004 *Triangulation. Eine Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Glaser, Barney G. und Anselm L. Strauss
1998 *Grounded Theory. Strategien qualitativer Forschung*. Bern etc.: Hans Huber.
- Grießhaber, Wilhelm
2006 Testen nichtdeutschsprachiger Kinder bei der Einschulung mit dem Verfahren der Profilanalyse – Konzeption und praktische Erfahrungen. In: Bernt Ahrenholz und Ernst Apeltauer (Hg.), *Zweitspracherwerb und curriculare Dimensionen. Empirische Untersuchungen zum Deutschlernen in Kindergarten und Grundschule*, 73–90. Tübingen: Stauffenburg.
- Grotjahn, Rüdiger
1987 On the methodological basis of introspective methods. In: Claus Faerch und Gabriele Kasper (Hg.), *Introspection in Second Language Research*, 54–81. Clevedon/Philadelphia: Multilingual Matters.
- Grotjahn, Rüdiger
2000 Testtheorie. Grundzüge und Anwendungen in der Praxis. In: Armin Wolff und Harald Tanzer (Hg.), *Sprache – Kultur – Politik*, 304–341. Universität Regensburg: FaDaF.

- Grotjahn, Rüdiger
2003 Konzepte für die Erforschung des Lehrens und Lernens fremder Sprachen. Forschungsmethodologischer Überblick. In: Karl-Richard Bausch, Herbert Christ und Hans-Jürgen Krumm (Hg.), *Handbuch Fremdsprachenunterricht*. 4., vollst. neu bearb. Aufl., 493–499. Tübingen/Basel: Francke.
- Grotjahn, Rüdiger
2005 Subjektmodelle: Implikationen für die Theoriebildung und Forschungsmethodologie der Sprachlehr- und Sprachlernforschung. *Zeitschrift für Fremdsprachenforschung* 16: 23–56.
- Grotjahn, Rüdiger
2006 Zur Methodologie der Fremdsprachenerwerbsforschung. In: Peter Scherfer und Dieter Wolff (Hg.), *Vom Lehren und Lernen fremder Sprachen. Eine vorläufige Bestandsaufnahme*, 247–270. Frankfurt a. M. etc.: Lang.
- Hatch, Evelyn und Anne Lazaraton
1991 *The Research Manual. Design and Statistics for Applied Linguistics*. New York: Newbury House.
- Heine, Lena
2005 Lautes Denken als Forschungsinstrument in der Fremdsprachenforschung. *Zeitschrift für Fremdsprachenforschung* 16: 163–185.
- Helfferrich, Cornelia
2004 *Die Qualität qualitativer Daten. Manual für die Durchführung qualitativer Interviews*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Henrici, Gert
1995 *Spracherwerb durch Interaktion? Eine Einführung in die fremdsprachenerwerbsspezifische Diskursanalyse*. Baltmannsweiler: Schneider.
- Kelle, Udo
1996 Die Bedeutung theoretischen Vorwissens in der Methodologie der Grounded Theory. In: Rainer Strobl und Andreas Böttger (Hg.), *Wahre Geschichten? Zu Theorie und Praxis qualitativer Interviews*, 23–47. Baden-Baden: Nomos.
- Krumm, Hans-Jürgen
2001 Unterrichtsbeobachtung und Unterrichtsanalyse. In: Gerhard Helbig, Lutz Götze, Gert Henrici und Hans-Jürgen Krumm (Hg.), *Deutsch als Fremdsprache. Ein internationales Handbuch*, 1139–1150. Berlin/New York: de Gruyter.
- Larsen-Freeman, Diane und Michael Long
1991 *An Introduction to Second Language Acquisition Research*. London/New York: Longman.
- Lienert, Gustav A. und Ulrich Raatz
1998 *Testaufbau und Testanalyse*. 6. Aufl. Weinheim: Beltz PsychologieVerlagsUnion.
- Mackey, Alison und Susan M. Gass
2005 *Second Language Research. Methodology and Design*. Mahwah, NJ/London: Erlbaum.
- Mayring, Philipp
2002 *Einführung in die qualitative Sozialforschung. Eine Anleitung zu qualitativem Denken*. 5. Aufl. Weinheim/Basel: Beltz.
- Mayring, Philipp
2008 *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. 10. Aufl. Weinheim etc.: Beltz.
- McDonough, Jo und Steven McDonough
1997 *Research Methods for English Language Teachers*. London: Arnold.
- McNamara, Tim
2000 *Language Testing*. Oxford: Oxford University Press.
- Nunan, David
1992 *Research Methods in Language Learning*. Cambridge: Cambridge University Press.
- Porst, Rolf
1996 Fragebogenherstellung. In: Hans Goebel, Peter H. Nelde, Zdenek Stry and Wolfgang Wölck (Hg.), *Kontaktlinguistik. Ein internationales Handbuch zeitgenössischer Forschung*. Band 1, 737–744. Berlin/New York: de Gruyter.

- Porst, Rolf
2008 *Fragebogen. Ein Arbeitsbuch*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Rasch, Björn, Malte Friese, Wilhelm Hofmann und Ewald Naumann
2006 *Quantitative Methoden 1*. 2. Aufl. Heidelberg: Springer.
- Richards, Keith
2009 Trends in qualitative research in language teaching since 2000. *Language Teaching* 42: 147–180.
- Riemer, Claudia
2007 DaF/DaZ und empirische Forschung: wechselnde Herausforderungen. *Info DaF* 34: 445–459.
- Schnell, Rainer, Paul B. Hill und Elke Esser
2005 *Methoden der empirischen Sozialforschung*. 7., völlig überarb. und erw. Aufl. München: Oldenbourg.
- Seliger, Herbert und Elana Shohamy
1989 *Second Language Research Methods*. Oxford: Oxford University Press.
- Silverman, David
1993 *Interpreting Qualitative Data. Methods for Analysing Talk, Text and Interaction*. London etc.: Sage.
- Steinke, Ines
1999 *Kriterien qualitativer Forschung. Ansätze zur Bewertung qualitativ-empirischer Sozialforschung*. Einheim/München: Juventa.
- Strauss, Anselm und Juliet Corbin
1996 *Grounded Theory. Grundlagen qualitativer Sozialforschung*. Weinheim: Beltz Psychologie VerlagsUnion.

Claudia Riemer, Bielefeld (Deutschland)
Julia Settinieri, Bielefeld (Deutschland)

86. Erstspracherwerb

1. Einleitung
2. Meilensteine des Erstspracherwerbs
3. Individuelle Unterscheide im Spracherwerb
4. Spracherwerb im Schulalter
5. Theorien des Spracherwerbes
6. Literatur in Auswahl

1. Einleitung

Kinder erleben von Geburt an die Bemühungen ihres sozialen Umfeldes um sprachliche Verständigung. In den ersten Lebensjahren lernen Kinder den Wortschatz und die Grundregeln der Syntax und Pragmatik einer oder mehrerer Sprachen. Sie können über ihr sprachliches Wissen in der Regel nicht explizit Auskunft geben, deshalb sind besondere methodische Ansätze erforderlich, um den Spracherwerb systematisch zu beobachten.