

Perspektiven der korpusbasierten Lernalersprachenanalyse aus fremdsprachendidaktischer Sicht

Anka Bergmann (Berlin)

В изучении методики и дидактики преподавания, а также в лингвистике существует традиция изучения и анализа ошибок в обучении иностранным языкам. Создание учебных корпусов и анализ речевой продукции учеников методами корпусной лингвистики открывает новые перспективы в изучении результатов школьного преподавания русского языка. Решение актуальных проблем в сфере диагностики ошибок и методах их устранения требует принятия во внимание современных материалов, а также их систематического анализа и оценки главным образом в контексте многоязычия, а также би-/полилингвальной дидактики. В данной статье будет представлен проект учебного корпуса, составленного на основе работ учащихся немецких школ и университетов.

1. Einführung

Lernerkorpora sind eine wichtige Ressource der Lernalersprachenanalyse in der Gegenwart. Sie ermöglichen nicht nur eine systematische Fehleranalyse, sondern erlauben eine darüber hinausgehende empirisch fundierte Beschreibung eines bestimmten Sprachstandes sowie den Vergleich verschiedener Lernalersprachen (in Bezug auf die L1 sowie in Bezug auf das Sprachniveau), bei entsprechender Architektur außerdem die Korrelation mit bestimmten Lernermerkmalen (z. B. Muttersprache, Spracherwerbsfolge usw.).

In diesem Beitrag wird das Projekt *Dokumentation und Analyse von Lernalersprache im Korpus* (DALeKo) vorgestellt, das auf eine Initiative des Arbeitskreises

Fremdsprachendidaktik (Anglistik, Romanistik, Slawistik) der Humboldt-Universität zu Berlin zurückgeht¹. Gegenstand des Projektes waren Entwicklung und Aufbau der Grundstruktur eines elektronischen Lernerkorpus, das Lernertexte in den Sprachen Englisch, Französisch, Russisch und Spanisch inkl. Metadaten zu den LernerInnen und zu den Bedingungen der Sprachproduktion enthält. Als Grundlage wurden schriftliche Klausuren und Lernerdaten aus Abiturjahrgängen in den Sprachen Englisch, Französisch, Russisch und Spanisch an Berliner Schulen erhoben. Aus fremdsprachendidaktischer Perspektive interessierte die InitiatorInnen² die Untersuchung von Lernalternativen unter den Bedingungen des institutionalisierten Erwerbs in der Schule. Die Idee war, parallel für mehrere Schulfremdsprachen Daten zu erheben und damit auch die Möglichkeit einer vergleichenden Perspektive zu haben. Derzeit wird der russische Teil des Korpus weiterentwickelt, dessen erster Teil bereits online durchsuchbar ist. In dem begrenzten Umfang des Projekts³ ging es zunächst darum, eine Struktur für das Korpus zu entwickeln, Texte zu erheben und zu diskutieren, welche Forschungsperspektiven ein solches Instrument für die Fremdsprachendidaktik bieten kann, insb. in welcher Weise Ergebnisse und Prozesse des Sprachenlernens in einem Kontext institutionalisierten Lernens betrachtet werden können. Fremdsprachenlernen in der Schule unterliegt besonderen Rahmenbedingungen und ist an tradierte und z. T. standardisierte Aufgabenformate gebunden, insofern handelt es sich bei den Lernertexten nicht um Beispiele spontaner freier Sprachproduktion. Für das Vorhaben gab es an der Humboldt-Universität mit dem Lernerkorpus FALCO (Fehlerannotiertes Lernerkorpus für Deutsch als Fremdsprache) ein

1 Eine Kurzdarstellung des Vorhabens und aktuelle Informationen zur Verfügbarkeit sind auf der Website des Projekts „Dokumentation und Analyse von Lernalternativen“ zu finden (<https://tinyurl.com/yc88mqkn> [02.12.2017]).

2 Prof. Anka Bergmann (Fachdidaktik Russisch), Prof. Stephan Breidbach (Fachdidaktik Englisch), Prof. Lutz Küster (Romanische Fachdidaktik).

3 Im Rahmen einer Projektförderung des Medienzentrums der Humboldt-Universität standen für ein Jahr drei studentische Hilfskräfte zur Verfügung, in der Folge gab es noch zwei kürzere Beschäftigungszeiträume für je eine Person. Da die Erarbeitung der Datenschutzrichtlinien und das Genehmigungsverfahren durch den Berliner Senat sehr viel Zeit in Anspruch nahmen, waren die Möglichkeiten der Auswertung der Texte sehr begrenzt.

konkretes Vorbild und in der Computerlinguistik eine ausgezeichnete Expertise, die für uns ein wertvoller Ansprechpartner war⁴.

2. Relevanz des Vorhabens

In Einführungen in die Korpuslinguistik wird darauf hingewiesen, dass Korpora als Grundlage für computergestützte Analysen sowohl für den Fremdsprachenunterricht als auch für die Fremdsprachenerwerbsforschung relevant sein können (z. B. Lemnitzer & Zinsmeister 2006, Nesselhauf 2004). Wie diese Nutzung aussehen kann, ist aber noch nicht ausreichend breit entwickelt oder thematisiert. Die Erstellung eines elektronischen Lernerkorpus soll in erster Linie darauf gerichtet sein, Lernaltersprache in größerem Umfang systematisch und vergleichend zu untersuchen. Für die Linguistik ist dies eine Möglichkeit, implizite Regeln des Sprachgebrauchs zu untersuchen, die erst durch Abweichungen z. B. in der Lernaltersprache deutlich werden (Rakhilina 2016).

Aus fremdsprachendidaktischer Perspektive sind insb. zwei Richtungen relevant: Zum einen soll ein Korpus fundierte Erkenntnisse über die Entwicklung von Lernaltersprache als Basis für die Auswahl und Begründung von Lehrverfahren, die Entwicklung von Lernaufgaben, die Gestaltung von Lehr-/Lernmaterialien, Nachschlagewerken sowie für die Bewertung liefern. Das heißt, es muss Erkenntnisse über häufige bzw. typische Fehler geben und sollte darüber hinaus auch Rückschlüsse zulassen auf Lernaltersprache, soweit sie sich bspw. in Form von Umschreibungen, Auslassungen etc. aus Lernaltersprachen erschließen lassen. Wir brauchen mehr und vor allem aktuelle Kenntnisse der typischen Entwicklung der Lernaltersprache auf bestimmten Niveaustufen. Für das Russische liegt lediglich eine bereits vor fast 25 Jahren erschienene Fehlersammlung vor (Borgwardt & Walter 1994). Seitdem haben sich sowohl der schulische Fremdsprachenunterricht, die Materialien als auch die LernerInnen gravierend verändert. Für alle Sprachenfächer ist die zunehmende Heterogenität der LernerInnen kennzeich-

4 Prof. Anke Lüdeling und ihren Mitarbeitern, insb. Mark Rezniczek, sind wir zu großem Dank für geduldige Antworten auf unsere Fragen und konkrete Unterstützung in technischen Problemen verpflichtet.

nend, es gibt infolge gesellschaftlicher Veränderungen viele mehrsprachige LernerInnen, die z. T. bereits mit Kenntnissen in mehreren Sprachen in die Schule kommen und auf dieser Basis weitere (Fremd-)Sprachen lernen. Auch der Beginn des schulischen Sprachenlernens ist vorgezogen, spätestens in Klasse 3 der Primarstufe beginnen heute alle SchülerInnen mit dem Englischlernen. Für den Russischunterricht sind die unterschiedlichen Lernvoraussetzungen von HerkunftssprecherInnen und FremdsprachenlernerInnen eine besondere Herausforderung (Bergmann 2016). Es gibt aber auch Veränderungen in der Zielsprache, die sich auf den Sprachunterricht auswirken. Seit 2008 ist eine neue Generation von Lehrbüchern auf dem Markt, die sich an den Anforderungen der Niveaustufen des Gemeinsamen europäischen Referenzrahmens für Sprachen (GeR) orientieren. Die Auswirkungen all dieser Veränderungen auf Lernprozesse und die Entwicklung der Lerner Sprache im schulischen Fremdsprachenunterricht sind bislang nur punktuell untersucht. Im Zuge der Standardisierung gelten für alle Fremdsprachen die gleichen Anforderungsbereiche (AFB) und Aufgabenarten. Damit könnte ein Lernerkorpus mit Daten aus dem Erwerb unterschiedlicher Sprachen die Vergleichsbasis für einzelsprachspezifische Erwerbsbesonderheiten auf unterschiedlichen Niveaustufen liefern und auch Schlüsse für Fragestellungen der Mehrsprachigkeitsdidaktik zulassen.

Eine zweite Perspektive für die Fremdsprachendidaktik ist die universitäre Lehre. Aktuelle authentische Lernerdaten sind maßgeblich für die Lehramtsausbildung. Im Zuge der Professionalisierung der LehrerInnenbildung wird vermehrt Aufmerksamkeit auf die Ausbildung von Diagnosekompetenz bei angehenden Lehrkräften gerichtet. In den fachdidaktischen Lehrveranstaltungen der Masterphase sollen die Studierenden zur Analyse von Schülerleistungen und zur darauf basierenden Einschätzung der Lernvoraussetzungen und Bewertung der Lernergebnisse befähigt werden. Bislang liegen dafür aber nur wenige, zudem unsystematisch gesammelte Daten vor, die keine dauerhaft qualitativ hochwertige Durchführung der Lehrveranstaltungen ermöglichen.

In der Praxis ist häufig zu beobachten, dass sich Lehrkräfte bei der Bewertung von Schülerleistungen ungeachtet der deklarierten Priorität des kommunikativen Erfolgs immer noch stark an formalen Kriterien orientieren. Dabei haben Lehr-

kräfte und Lehramtsstudierende erfahrungsgemäß eigene und z. T. fest gefügte Vorstellungen von den Bereichen, die LernerInnen besondere Schwierigkeiten bereiten. Interessanterweise unterscheiden sich nach meiner Erfahrung diese Vorstellungen bei denjenigen, für die Russisch Herkunftssprache ist, oder die es selbst als Fremdsprache gelernt haben. Angesichts der Tatsache, dass es zu diesem Thema in den letzten zwei Jahrzehnten keine nennenswerte Forschung gibt, folgen Schwerpunkte in Curricula, in Lehr- und Lernmaterialien und in der Gestaltung des Lehr-/Lernprozesses zum einen traditionellen Mustern und resultieren zum anderen aus alltäglichem Erfahrungswissen und subjektiven Theorien der Lehrkräfte. Damit in der Unterrichtspraxis eine differenzierte Förderung stattfinden kann, die an den Stärken der SchülerInnen ansetzt und gezielt die Bereiche fördert, die für sie schwierig sind, müssen Lehrkräfte in der Lage sein, die Eingangsvoraussetzungen der LernerInnen einzuschätzen. Auch für Lehrpersonen, die schon lange in der Schule tätig sind, werden Schulungen und Fortbildungsmaterialien benötigt. Diagnoseinstrumente, die im Schulalltag handhabbar sind und von LehrerInnen selbstständig als Beobachtungshilfe genutzt werden können (wie etwa im DaZ-Bereich die Profilanalyse von Griefhaber 2013), müssen für das Russische erst entwickelt werden.

3. Theoretisch-konzeptuelle Grundlagen und methodologische Vorüberlegungen

Bevor ich auf Details des DALeKo-Projekts eingehe, möchte ich einige grundsätzliche Aspekte besprechen, die wichtige Vorüberlegungen für ein Lernerkorpusprojekt darstellen:

Ein Korpus ist zunächst eine Sammlung von Sprachbeispielen bzw. Texten, die nach expliziten linguistischen Kriterien ausgewählt und angeordnet sind (vgl. EAGLES 1996). Dies können mündliche oder schriftliche Sprachbeispiele sein. Im Unterschied zu herkömmlichen Textsammlungen stellt ein Korpus in der Regel eine digitalisierte Textsammlung dar, die formale sprachliche Markierungen enthält und mit entsprechenden Werkzeugen durchsuchbar ist. Aufbau,

Auszeichnung und Auswertung von Korpora sind Gegenstand der Korpuslinguistik⁵ (Lüdeling & Walter 2009: 1). In deren Kontext entwickelt sich seit den 1980er Jahren die Beschäftigung mit Lernerkorpora. Natürlich gibt es Analysen von Lernaltersprache v. a. im Sinne von Fehlersammlungen und Fehlerklassifikationen bereits viel länger. Im Vergleich zu nichtdigitalen Fehlersammlungen, deren begrenzter Umfang und Zusammenstellung der Beispiele direkt von der Fragestellung und den theoretischen Vorannahmen und/oder einer konkreten Lehrsituation bestimmt ist, sind Lernerkorpora für eine Vielzahl von Fragestellungen offen. Der Zugriff auf große Datenmengen und die automatisierte Abfrage gewähren neue Möglichkeiten der Analyse. Im Zuge der Annotation können viele unterschiedliche Informationen zu den Texten (Textsorte, Erscheinungsdatum), zum/zur TextproduzentIn sowie zu den Bedingungen der Textproduktion aufgenommen werden, nach denen gesucht werden kann (zu Annotation vgl. Lemnitzer & Zinsmeister 2006: 60–100).

Dafür müssen Korpora bestimmten Designkriterien genügen. Sie enthalten authentische Textdaten (eine Fehlersammlung ist kein Lernerkorpus) und dokumentieren die Bedingungen der Textproduktion, vgl. folgende Definition: “Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardised and homogenous way and documented as to their origin and purpose” (Granger 2002: 7). Die Lernerkorpusforschung setzt sich zum Ziel, die Beschreibung von Lernaltersprache zu verbessern, um Lernprozesse besser zu verstehen und die Effizienz des Sprachunterrichts zu erhöhen (ebd.). Dafür wird angestrebt, durch große Datenmengen und ein entsprechendes Design, Repräsentativität zu erreichen. Die elektronische Verfügbarkeit der Daten ermöglicht eine rasche strukturierte Analyse der Daten, die zudem wiederholt werden kann. Die Ergebnisse sind dadurch allgemein nachvollziehbar. Eine Analyse mit korpuslinguistischen Methoden kann zum einen Informationen über die Häufigkeit linguistischer Einheiten (Wortbestandteile, Wörter, grammatische Strukturen) liefern und zum anderen über die Kontexte, in denen diese

5 Deutschsprachige Einführungen in die Korpuslinguistik sind z. B. Lemnitzer & Zinsmeister 2006, Perkuhn, Keibel & Kupietz 2012, Scherer 2006.

aufzutreten: Kontextfaktoren können systematisch ausgewertet werden. Dies geschieht in der Regel in Form von Konkordanzen. Die Daten im DALeKo-Projekt sind zwar bislang nicht umfangreich, sodass Häufigkeitsuntersuchungen lediglich eine sehr begrenzte Aussagekraft haben, aber eine Kontextanalyse ist auch für eine begrenzte Lernergruppe von Interesse. Sind die Korpora öffentlich zugänglich, können vielfältige Fragestellungen am gleichen Material untersucht werden (vgl. Lüdeling & Walter 2009).

Beim Aufbau eines Lernerkorpus war für uns in erster Linie die Frage zu diskutieren, welche Texte aufgenommen werden sollen. Die Lernerkorpusforschung geht davon aus, dass bei der Kompilation das Prinzip der Authentizität gilt, d. h. aufgenommen werden Texte, die in natürlichen, authentischen Kommunikationssituationen verfasst wurden (Mukherjee 2002). Das klassische Untersuchungsdesign der L2-Forschung sieht vor, dass relevante Daten nur solche sind, bei denen der/die L2-Lernende versucht, bedeutungsvolle Äußerungen in der Zielsprache zu machen (Selinker 1972). Dies folgt aus der Grundannahme, dass Rückschlüsse auf die Zweitsprachkompetenz der Lernenden nur über die Beobachtung und Beschreibung lernaltersprachlicher Performanzphänomene möglich sind (Ellis & Barkhuizen 2005). Aus fremdsprachendidaktischer Perspektive zeigt sich aber, dass der Begriff ‚bedeutungsvoll‘ (*meaningful*) nicht unproblematisch ist. Vor allem greift er zu kurz, wenn er auf den kommunikativen Gehalt sprachlicher Äußerungen beschränkt ist und dieser Gehalt dabei ausschließlich an den Bedürfnissen außerunterrichtlicher Kommunikation gemessen wird. Streng genommen würden damit sehr viele Äußerungen aus unterrichtlichen Konstellationen nicht als Datenquelle in Frage kommen, da sie bestimmte sprachliche Strukturen aus unterrichtlichen Überlegungen heraus erfordern und diese speziell eingeübt werden. Der/die LernerIn wählt sprachliche Mittel unter dem Einfluss verschiedener Faktoren aus, nicht allein danach, ob sie für den kommunikativen Erfolg bedeutsam sind. Aus der Verwendung einer sprachlichen Form oder Struktur im institutionalisierten Sprachunterricht kann man letztlich nicht darauf schließen, ob Lernende diese Strukturen auch außerhalb des Unterrichts benutzen würden oder ob sie nicht auf anderem Wege ebenso kommunikativ angemessen reagieren könnten. Den Lernenden ist ja durchaus bewusst, dass sie sich mit sprachlichen Strukturen nicht nur mit Blick auf deren kommunikative Leistungsfähigkeit aus-

einandersetzen, sondern auch mit der Absicht, sich diese anzueignen. Ihnen ist zudem bewusst, dass ihre zielsprachliche Performanz Gegenstand von Überprüfung ist oder werden kann und dass in diesem Zusammenhang bestimmte Strukturen erwartet werden und es nicht oder zumindest nicht nur um die erfolgreiche Bewältigung einer kommunikativen Situation geht (Königs 2006). Für LernerInnen sind auch und gerade unterrichtliche Situationen und Prüfungssituationen authentisch und gegenwartsbezogen. In der fremdsprachendidaktischen Diskussion zur Aufgabenorientierung des gesteuerten Fremdsprachenlernens wird mit Blick auf die kommunikative Bedeutsamkeit der Begriff der Aufgabe verwendet, um eine Abgrenzung zu rein formorientierten und oftmals bedeutungsleeren Übungen vorzunehmen. Eine Aufgabe gilt als bedeutungsvoll, weil sie so angelegt sein muss, dass sie den/die LernerIn mit für ihn/sie relevanten kommunikativen Aufgabenstellungen konfrontiert (ebd.: 115) und ihn/sie zu einer Sprachproduktion motiviert, die sein/ihr sprachliches Handlungsvermögen abbildet.

Das didaktische Interesse bei der Analyse von Lernertexten bezieht sich auf die Frage, welche Anforderungen bestimmte Arten von Aufgabenstellungen an die LernerInnen stellen und wie die LernerInnen diese in einer unterrichtlichen Situation erfüllen. Eine typische Aufgabenstellung im institutionalisierten Lernen sowohl in der Schule als auch im universitären Sprachunterricht ist z. B. die zusammenfassende schriftliche Wiedergabe eines fremdsprachigen Textes (Bergmann & Ermakova 2017). Uns interessiert in diesem Fall, welche Anforderungen eine Aufgabe zur Produktion einer Textzusammenfassung (mit oder ohne eigene Stellungnahme zum Thema) an den Lernenden stellt, welche Art von Performanz sie initiiert und wie sich diese zu den Anforderungen authentischer außerunterrichtlicher Kommunikation verhält. Da schriftliche Textzusammenfassungen auch als Prüfungsformat verwendet werden, nimmt die Vorbereitung darauf im Unterricht viel Raum ein und kann leicht zu *Washback*-Effekten führen. Umso wichtiger ist es, die Aufgabenformate in Beziehung zu den Zielen eines kompetenzorientierten Unterrichts und den tatsächlichen Anforderungen kommunikativen Handelns in der Zielsprache zu setzen. Im Rahmen der korpuslinguistischen Analyse sind die Aufgabenstellung sowie Situationsfaktoren (z. B. Zeitlimit, Verwendung von Hilfsmitteln bei der Textproduktion u. ä.) und Informationen zu den Lernvoraussetzungen der TextproduzentInnen zu berücksichti-

gen, um die kommunikativen Anforderungen zu spezifizieren. Auf dieser Grundlage können dann zur Herstellung von Vergleichbarkeit ggf. Teilkorpora gebildet werden. Werden die Ausgangstexte mit aufgenommen, wird ein Vergleich von Ausgangsstrukturen und deren lernerseitiger Wiedergabe möglich (Wann werden Formulierungen übernommen? In welchem Umfang? Welche Modifizierungen gibt es?) Ein solcher Ansatz geht weiter, als sich auf die Ermittlung und Analyse fehlerhafter Äußerungen zu konzentrieren. Interessant ist gerade auch, was der/die Lernende schon beherrscht und wie er/sie bestimmte sprachliche Mittel zur Erreichung seiner/ihrer kommunikativen Zwecke einsetzt (das kann auch die Erfüllung der Aufgabenstellung sein). Dies führt über die klassische Fehlerklassifizierung hinaus und überwindet die in der Lehrpraxis immer noch vielfach zu beobachtende Begrenzung in Diagnose und Bewertung auf die morphologisch-syntaktische Ebene. Es wird darauf geblickt, wie die Sprachproduktion zu einem bestimmten Zeitpunkt im institutionalisierten Lernen aussieht. Nun ist Lernaltersprache in höchstem Maße individuell und verändert sich fortlaufend unter dem Einfluss unterschiedlicher Faktoren. Das macht die Verallgemeinerung von Untersuchungsergebnissen schwierig. Es ist aber anzunehmen, dass es ungeachtet der Variabilität und Dynamizität invariante Bereiche in der Lernaltersprache gibt, die für die Lernaltersentwicklung in bestimmten Gruppen und auf bestimmten Stufen kennzeichnend sind, und dass sprachliche Oberflächenmerkmale in Lernaltersertexten als Indikatoren für die Einschätzung des erreichten Lernstandes eines/einer Lernenden genutzt werden können. Auf solchen Vorannahmen beruhen z. B. die von Grieffhaber entwickelte Profilanalyse für Deutsch als Zweitsprache (Grieffhaber 2006, 2013), das Zürcher Textanalyseraster (Nussbaumer 1996) oder die Beschreibung von Grammatikerwerbsfolgen auf der Basis deutschsprachiger Aufsätze französischsprachiger SchülerInnen in der Schweiz (Diehl et al. 2000). Für fremdsprachendidaktische Entscheidungen ist relevant, wie die Sprachproduktion auf bestimmten Niveaustufen aussieht, die für das institutionalisierte Fremdsprachenlernen angenommen werden, was etwa auf Abiturniveau in Grund- oder Leistungskursen erwartbar ist, was wirklich erreicht wird, welche sprachlichen Anforderungen traditionell im Sprachunterricht verwendete Textsorten an die LernerInnen stellen. Für eine empirisch fundierte Beschreibung des tatsächlichen Sprachgebrauchs der LernerInnen können Methoden der Korpuslinguistik folg-

lich sehr nützlich sein. Diese Fragestellungen sind im Forschungsfeld angesiedelt, wie weit LernerInnen im institutionalisierten Fremdsprachenlernen gelangen können und wie dies korreliert mit Faktoren wie Lernbeginn (L3 im Vergleich zu L2), Alter, Motivation, Vorkenntnisse, Sprachlerninteresse u. ä.

4. Das DALeKo-Korpus

4.1. Lernertexte

Für das DALeKo-Korpus (Dokumentation und Analyse von Lernersprache im Korpus) wurden Abiturvorklausuren gesammelt, die im Rahmen des regulären Fremdsprachenunterrichts an Berliner Schulen geschrieben wurden (11., 12. und 13. Klassen). Um ein differenziertes Spektrum an Lernertexten zu erfassen, sind pro Sprache (Englisch, Französisch, Spanisch, Russisch) jeweils ein Grundkurs (GK) und ein Leistungskurs (LK) vertreten. Für alle Sprachen wurden insgesamt ca. 750 Einzeltexte aus sieben Berliner Schulen erhoben. Dazu wurden in einem Fragebogen jeweils Informationen zur Sprachbiografie und zum Sprachenlernen erfasst. Die Probeklausuren wurden zu unterschiedlichen Themen im Rahmen der Semesterschwerpunkte für den Fremdsprachenunterricht in der Sekundarstufe II geschrieben und stimmen hinsichtlich Aufgabentyp und entsprechender Bearbeitungszeit weitgehend mit den Prüfungsbedingungen von Abiturklausuren überein. Entsprechend den Einheitlichen Anforderungen für die Abiturprüfung (EPA Russisch 2004) sind Einzeltexte drei Anforderungsbereichen zuzuordnen:

1. Reproduktion und Textverstehen,
2. Reorganisation und Analyse,
3. Werten und Gestalten.

Für den AFB 1 werden in der Regel Textzusammenfassungen geschrieben. Die Aufgabenstellung lautet z. B. *Резюмируйте данный текст* oder *Резюмируйте главные мысли текста*. Diese Texte werden im Korpus als *summary* gekennzeichnet. Darüber hinaus gibt es entweder eine oder zwei Aufgaben zur Textanalyse,

zur Argumentation oder zum Ausdruck der eigenen Meinung auf Basis der Ausgangstexte. Nicht in jedem Fall gibt es getrennte Aufgaben für AFB 2 und 3. Diese werden als *analysis* gekennzeichnet. Der zeitliche Rahmen für die Klausuren ist dann je nach Umfang entweder 180 oder 240 min. Die Einzeltexte haben im Schnitt einen Umfang von 190 Wörtern. Die Wortanzahl der Textvorlagen umfasst ca. 300–700 im GK und ca. 450–900 Wörter im LK. In der Regel wird der Einsatz eines einsprachigen Wörterbuchs zugelassen. Eine Mindest-Wortanzahl wird bei der Textproduktion nicht explizit verlangt, in einem Fall ist zu den Aufgabenstellungen intern vermerkt, dass eine Mindestwortzahl von insgesamt 350 Wörtern (für alle Aufgaben der Klausur) erwartet wird. Für den AFB 1 Reproduktion (Textzusammenfassung) ist einmal eine Obergrenze von max. 120 Wörtern gesetzt. Die Texte lagen zunächst handschriftlich vor, sie wurden von jeweils zwei MitarbeiterInnen abgetippt und die Versionen in einer Endversion zusammengeführt. Dies ist nach den Erfahrungen aus anderen Lernerkorpora notwendig, um nicht Fehler bei der Übertragung zu korrigieren, und hat sich bei unserer Arbeit bestätigt. Aus allen Texten wurde ein Pilot-Korpus erstellt, das nach Sprache und Leistungsstufe (GK und LK) geordnet ist. Pro Sprache gibt es je 10 GK und 10 LK-Texte aus einer Schule (i. d. R. in einer Klasse erhoben), die mit dem Programm *Tree Tagger* automatisch nach Wortarten annotiert wurden. Für die unterschiedlichen Sprachen wurden unterschiedliche *Tagsets* eingesetzt.⁶ Auf der Basis dieser Annotation lässt sich bei automatischen Suchabfragen klären, welche Strukturen überhaupt in den Texten vorkommen, bezogen auf den/die einzelne/n LernerIn oder auf alle Texte eines (Teil-)Korpus.

4.2. Fehler und lernalterssprachliche Abweichungen

Fehler bzw. Abweichungen von der zielsprachlichen Norm in den Lernertexten können nicht automatisch markiert werden. Ein methodologisches Problem ist, dass jede Korrektur eine Hypothese darüber voraussetzt, was als korrekt angesehen wird, vor allem auf der Ebene des Satzes oder Textes ist dies aber nicht immer eindeutig zu entscheiden: Sprache ist selbst im Rahmen der standard-

⁶ Für das Russische <https://tinyurl.com/yb7qb8ss> [02.12.2017].

sprachlichen Norm in gewissem Maße variabel und individuell und ein Gedanke kann auf unterschiedliche Art ausgedrückt werden. Bei einer Fehlerklassifikation muss sich der/die AnnotatorIn außerdem in jedem Fall entscheiden, auf welcher sprachlichen Ebene der Fehler auftritt oder was seine Ursache sein könnte. Dies ist nicht immer eindeutig und kann individuell zu unterschiedlichen Einschätzungen führen. Wenn im Russischen z. B. eine Endung falsch geschrieben ist, so kann dies darauf zurückzuführen sein, dass der/die LernerIn die Endung nicht kennt, was oft bei FremdsprachenlernerInnen der Fall ist. Dies wäre als morphologischer Fehler einzuordnen. Oder der/die LernerIn kennt die Endung zwar mündlich in einer reduziert ausgesprochenen Form, hat aber nicht gelernt, wie sie geschrieben wird. Dies wäre dann ein orthographischer Fehler und eher für HerkunftssprecherInnen typisch (s. die Beiträge von Brüggemann und Ermakova in diesem Band). Eine verlässliche Fehlerkategorisierung durchzuführen und Lernertexte entsprechend zu annotieren, ist daher mit einem erheblichen Arbeitsaufwand verbunden. Für das *Russian Learner Corpus*⁷ (RLC) wurde in Moskau eine Fehlerklassifikation entwickelt, die im Zuge eines *annotator interagreement experiment* evaluiert wurde und laufend überarbeitet wird (vgl. Rakhilina 2016).

Anstelle einer Fehlerklassifikation kann ein Lernerkorpus mit dem Konzept der Zielhypothese arbeiten, wie dies bei FALKO gehandhabt wird. Die Zielhypothese stellt in einer ersten Version (ZH 1) eine grammatische Variante des Ausgangstextes dar: “reconstruction of those utterances in the target language” (Ellis 1994: 54). Es werden orthographische, morphologische und syntaktische Abweichungen von der standardsprachlichen Norm korrigiert. Semantik bzw. Wortwahl, Angemessenheit, Schriftsprachlichkeit oder muttersprachliche Nähe spielen in der ZH 1 keine Rolle. Sie können in einer zweiten ZH berücksichtigt werden, die unter Einbeziehung lexikalischer, semantischer und pragmatischer Information eine nicht nur grammatisch korrekte, sondern funktional äquivalente Zielstruktur herstellt. Dies ist entsprechend aufwändiger und kann von uns nicht realisiert werden, ist aber eine lohnende Form für die Diskussion in fachdidaktischen und linguistischen Seminaren. Zuweilen sind mehrere ZH 1 möglich, jedoch ist festgelegt, dass Analyse und Korrektur ausgehend vom Verb stattfinden, sodass

7 <https://tinyurl.com/y9q8bjrj> [02.12.2017].

die Argumentstruktur an das geschriebene Verb angepasst wird, selbst wenn die geschriebene Argumentstruktur unter Auswahl eines geeigneteren Verbs korrekt gewesen wäre. Die Zielhypothese muss manuell in eine *Excel*-Tabelle eingetragen werden, dann kann sie wiederum automatisch annotiert und durchsucht werden und es lassen sich Abweichungen zwischen Lernertext und Zielhypothese finden. Das Konzept der Zielhypothese bedeutet nicht die Rekonstruktion der Lernerintention (vgl. Lemnitzer & Zinsmeister 2006: 86). Selbst wenn der Lernertext eine Nacherzählung oder Zusammenfassung eines vorliegenden Ausgangstextes darstellt, mit dem man vergleichen kann, ist es kaum möglich, die Lernerintention hypothetisch nachzubilden. Der Gedanke des Ausgangstextes kann von dem/der Lernenden sprachlich in unterschiedlicher Formulierung wiedergegeben werden. Dies hängt nicht nur von dessen/deren sprachlichem Vermögen ab; es ist vielmehr ein didaktischer Faktor, dass oftmals gerade das Wiedergeben des Textes ‚mit eigenen Worten‘ verlangt wird, sodass der/die LernerIn bestrebt sein müsste, andere Konstruktionen als im Ausgangstext zu benutzen. Die Reformulierung eines Textes stellt insofern sehr hohe Anforderungen an SprachbenutzerInnen: Sie müssen den Ausgangstext verstehen, die Informationen gewichten und strukturieren und schließlich einen eigenen Text sprachlich angemessen formulieren. Dieser soll dann erheblich kürzer als der Ausgangstext sein. Insgesamt ist dies eine Aufgabe, die auch im muttersprachlichen Bereich erlernt und geübt werden muss. Manchmal stellt sich die Frage, ob der/die Lernende den Ausgangstext ausreichend verstanden hat, z. B. wenn es Übernahmen aus dem Ausgangstext gibt, die sich durch deutlich weniger Fehler auszeichnen als andere Textstellen. Aus Textzusammenfassungen lässt sich aufgrund dieser Spezifik nicht in gleicher Weise auf die sprachliche Kompetenz des/der Lerner/schließen wie bei freien Schreibproduktionen.

Zunächst wurde im DALeKo-Korpus je ein Text aus dem GK in jeder Sprache beispielhaft mit einer Zielhypothese versehen. Im Rahmen der aktuellen Erweiterung des russischen Korpus liegen jetzt 117 Schülertexte vor, die getaggt, mit Zielhypothesen versehen und mit Metadaten verknüpft sind. Diese werden bei FALKO ‚gehostet‘, d. h., dass sie dort mit ANNIS durchsuchbar sind. Ein Zugang wird auf Anfrage gewährt, dies entspricht den Datenschutzrichtlinien, die im Projekt für den Umgang mit Schülertexten erarbeitet wurden. Außer-

dem können die Metadatendokumentationen und Volltexte auf Basis einer Nutzungsvereinbarung ausgegeben werden. Genauere Informationen sind auf der Homepage des Projekts⁸ zu finden. In der Perspektive sollen weitere Texte von Studierenden im Rahmen von linguistischen und fachdidaktischen Seminaren mit Zielhypothesen versehen werden. Dies erfordert linguistisches Wissen, schult die Analysefähigkeit und kann die Reflexion über das eigene Sprachenlernen anregen. Das Korpus soll in Zukunft um Texte aus dem universitären Sprachunterricht erweitert werden. Momentan liegen 39 studentische Texte, ebenfalls schriftliche Textzusammenfassungen mit Ausgangstexten, vor. Zu überlegen ist außerdem eine Erweiterung um Texte, die auf der Grundlage freier Aufgabenstellungen entstehen. Derzeit werden die Aufgabenstellungen und die Ausgangstexte für die Aufnahme ins Korpus erfasst.

4.3. Metadaten zu Lernermerkmalen

Die Entwicklung eines Metadatenmodells setzt Annahmen darüber voraus, welche Faktoren das Lernen im institutionellen Kontext beeinflussen und welche der Voraussetzungen, mit denen LernerInnen in den Unterricht der ersten, zweiten oder dritten Fremdsprache kommen, relevant für die sprachliche Leistung sein können. Mit der zunehmenden Diskussion von mehrsprachigkeitsdidaktischen Fragen und insb. mit der Untersuchung der Sprachkompetenzen und Sprach(lern-)einstellungen von mehrsprachigen SchülerInnen sind solche Faktoren in den letzten Jahren vermehrt ins Blickfeld gerückt. Diese Faktoren betreffen die individuellen Voraussetzungen der Lernenden, ihre Muttersprache(n), Besonderheiten der Spracherwerbssituation, des Lernwegs, die Einstellungen zum Sprachenlernen insgesamt, die sich nicht zuletzt aus der Wertschätzung der Sprachkompetenzen im gesellschaftlichen Kontext und in der sozialen Interaktion ergeben, aber auch Merkmale des alltäglichen Sprachgebrauchs. Während das RLC im Wesentlichen die Zuordnung als HerkunftssprecherIn oder FremdsprachenlernerIn (*author's language background*), die dominante Sprache (*author's dominant language*) sowie das Sprachkompetenzniveau (*author's proficiency in Russian*) als

8 <https://tinyurl.com/yc88mqkn> [02.12.2017].

Pflichtangaben aufnimmt, haben wir uns für einen umfangreichen Fragebogen entschieden, um Ansatzpunkte für unterschiedliche Fragestellungen zu bieten. Für das Russische ist die Unterscheidung von L1- und L2-SprecherInnen wichtig. Dem liegen Erkenntnisse der Spracherwerbsforschung (Gagarina 2017) und praktische Unterrichtserfahrungen zugrunde, dass diese beiden Lernergruppen sich in ihren Spracherwerbsverläufen unterscheiden und jeweils Spezifika aufweisen, die die Lernwege und -ergebnisse beeinflussen können. Da es aber auch bei den HerkunftssprecherInnen eine sehr große Varianz in den Sprachbiografien und in der Sprachkompetenz gibt (ebd.), ist es von Interesse, detailliertere Daten zum Lernen und zum Sprachgebrauch zu haben. Die Fragen beziehen sich auf folgende drei Schwerpunkte:

- Sprach(lern)biografie:
Muttersprache(n),
gelernte Sprachen,
Sprachlernerfahrung schulisch/außerschulisch
- Einstellungen:
zum Sprachenlernen,
empfundener Schwierigkeitsgrad der Zielsprache
- Sprachgebrauch:
zu Hause gesprochene Sprachen,
im Alltag für verschiedene Aktivitäten bevorzugte Sprachen (Lesen, Medienutzung, soziale Netzwerke),
Sprache in der Interaktion mit verschiedenen Personen (schematisch mit Pfeilen gekennzeichnet)

Alle Metadaten außer diesen Schemata wurden in *Excel*-Tabellen eingepflegt und können jeweils über den Dateititel der einzelnen Texte zugeordnet werden. Bei annotierten Texten in *Microsoft Excel* sind die Metadaten über einen weiteren Reiter direkt mit dem jeweiligen Text verknüpft.

5. Perspektiven

Die im DALeKo-Projekt angelegte Korpusstruktur ist sowohl quantitativ als auch qualitativ erweiterbar. Hinsichtlich der qualitativen Erweiterung sind weitere LernerInnen-texte aus verschiedenen Erwerbskontexten (gleiche oder andere Jahrgänge, immersive Vergleichsklassen und Europaschulen) von Interesse. Für eine reliable statistische Auswertung wäre eine quantitative Erweiterung notwendig, da bei einer so geringen Korpusgröße die Daten stark von individuellen und lerngruppenabhängigen Besonderheiten verzerrt werden. Aber bereits mit einer geringen Größe eines Korpus können deskriptive statistische Berechnungen realisiert werden, die z. B. die Häufigkeit der Verwendung bestimmter Lemmata, Konstruktionen oder grammatischer Formen, die Textlänge und das Auftreten von *chunks* zeigen. Es kann untersucht werden, welche Formen regelgerecht gebildet und kontextangemessen verwendet werden. Von Interesse sind darüber hinaus lexikalische Variation, Mittel der Textstrukturierung u. ä. Die ermittelten Besonderheiten lassen sich in Beziehung setzen zu Faktoren des Lernprozesses wie Lerdauer, Alter, Sprachlernerfahrung, Mehrsprachigkeit, Kurswahl (GK oder LK), Sprachumgebung, Selbstbild in Bezug auf das Sprachenlernen und Sprachverhalten im Alltag. Das Lernerkorpus soll in erster Linie ein Untersuchungs- und Arbeitsinstrument im Rahmen der Lehramtsausbildung sein. Wir wollen Aussagen darüber gewinnen, welche Texte SchülerInnen in der Sekundarstufe II im Russischunterricht produzieren bzw. wie die Lernaltersprache Russisch auf einer bestimmten Niveaustufe aussieht, welche Besonderheiten die Sprachproduktion von HerkunftssprecherInnen und FremdsprachenlernerInnen aufweist und wie diese mit den Anforderungen des schulischen Fremdsprachenunterrichts korrelieren und wie sie allgemein in bildungspolitischen Vorgaben formuliert sind. Auf dieser Basis können Diagnoseinstrumente und Materialien für Studierende und Lehrkräfte entwickelt werden, um sie zum einen für den individuellen Sprachentwicklungsprozess der SchülerInnen und zum anderen für die Problemfelder zu sensibilisieren, die auf einer bestimmten Entwicklungsstufe bzw. bei bestimmten Aufgabenstellungen auftreten. Die Vorbereitung der Texte aus dem Sprachunterricht für die Aufnahme ins Korpus soll Bestandteil der fachdidaktischen und

sprachlichen Ausbildung der Lehramtsstudierenden sein. Denkbar sind hier auch Kooperationen mit anderen slawistischen Instituten. Studentische Texte aus dem universitären Sprachunterricht werden wie bislang vorzugsweise in das Moskauer RLC weitergegeben, in dem Texte von RussischlernerInnen aus verschiedenen Sprachumgebungen repräsentiert sind und das in größerem Umfang mehr Untersuchungsmöglichkeiten, nicht aber die Perspektive auf den deutschen schulischen Kontext bietet.

Literaturverzeichnis

- Bergmann, A. & Ermakova, N. (2017). Преподавание русского языка в немецкой аудитории. In *Русский язык за рубежом. Спецвыпуск: Русистика Германии*. (41–45). Москва.
- Bergmann, A. (2016). Lernvoraussetzungen und Differenzierungsansätze im Russischunterricht: Einige Überlegungen für einen lernerorientierten schulischen Russischunterricht. In A. Bergmann (Hrsg.), *Kompetenzorientierung und Schüleraktivierung im Russischunterricht* (43–70). Frankfurt a. M.: Peter Lang.
- Borgwardt, U. & Walter, H. (1994). *Fehler-ABC: Deutsch-Russisch*. Stuttgart u. a.: Klett.
- Diehl, E. et al. (2000). *Grammatikunterricht: Alles für der Katz? Untersuchungen zum Zweitspracherwerb Deutsch*. Tübingen: Niemeyer.
- EAGLES (1996). *Preliminary Recommendations on Corpus Typology. EAGTCWG-CTYP/P*. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale. Verfügbar unter: <https://tinyurl.com/ybww8qjp> [02.12.2017].
- Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Ellis, R. & Barkhuizen, G. (2005). *Analysing Learner Language*. Oxford u. a.: Oxford University Press.
- Gagarina, N. (2017). Monolingualer und bilingualer Erstspracherwerb des Russischen. Ein Überblick. In K. Witzlack-Makarevich & N. Thiel, *Handbuch des Russischen in Deutschland* (393–410). Berlin: Frank & Timme.
- GeR=Trim, J., North, B. & Coste, D. (Hrsg.) (2001). *Gemeinsamer europäischer Referenzrahmen für Sprachen: Lernen, lehren, beurteilen*. Berlin: Langenscheidt.
- Granger, S. (2002). A Bird's-eye View of Learner Corpus Research. In S. Granger, J. Hung, St. Petch-Tyson (Hrsg.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* (3–33). Amsterdam: John Benjamins.

- Grießhaber, W. (2006). *Sprachstandsdiagnose im kindlichen Zweitspracherwerb: Funktional-pragmatische Fundierung der Profilanalyse*. Verfügbar unter: <https://tinyurl.com/yc5gj2t4> [02.12.2017].
- Grießhaber, W. (2013). *Die Profilanalyse für Deutsch als Diagnoseinstrument zur Sprachförderung*. Verfügbar unter: <https://tinyurl.com/y8e2k6lg> [02.12.2017].
- Königs, F. (2006). Aufgabenorientierung als Aufgabe. Überlegungen zu einem ‚neuen‘ Konzept für den Fremdsprachenunterricht. In K.-R. Bausch, E. Burwitz-Melzer, F. G. Königs & H.-J. Krumm (Hrsg.), *Aufgabenorientierung als Aufgabe. Arbeitspapiere der 26. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts* (115–122). Tübingen: Narr.
- Lemnitzer, L. & Zinsmeister, H. (2006). *Korpuslinguistik. Eine Einführung*. Tübingen: Narr.
- Lüdeling, A. & Walter, M. (2009). *Korpuslinguistik für Deutsch als Fremdsprache. Sprachvermittlung und Spracherwerbsforschung*. Verfügbar unter: <https://tinyurl.com/y924bvcz> [02.12.2017].
- Mukherjee, J. (2002). *Korpuslinguistik und Englischunterricht. Eine Einführung*. Frankfurt a. M.: Peter Lang.
- Nesselhauf, N. (2004). Learner Corpora and Their Potential for Language Teaching. In J. Sinclair (Hrsg.), *How to Use Corpora in Language Teaching* (125–152). Amsterdam: John Benjamins.
- Nussbaumer, M. (1996). Lernerorientierte Textanalyse – Eine Hilfe zum Textverfassen? In H. Feilke & P. Portmann (Hrsg.), *Schreiben im Umbruch. Schreibforschung und schulisches Schreiben* (96–112). Stuttgart: Klett.
- Perkuhn, R, Keibel, H. & Kupietz, M. (2012). *Korpuslinguistik*. Paderborn: Fink.
- Rakhilina, E. (2016) = Рахилина, Е. О новых инструментах описания русской грамматики: корпус ошибок. *Русский язык за рубежом*, 3, 20–25.
- Scherer, C. (2006). *Korpuslinguistik*. Heidelberg: Winter.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching (IRAL)*, 10, 31–54.