

Untersuchung von Textkorpora

1. Einleitung

Die Untersuchung von Textkorpora, d.h. Textsammlungen, die zielgerichtet und systematisch erstellt worden sind, gehört in der schreibdidaktischen Forschung zu den am häufigsten genutzten Methoden. Dies gilt insbesondere für die Schreibentwicklungsforschung im deutschsprachigen Raum, die von Beginn an mit Textkorpora gearbeitet hat. Ein bekanntes Beispiel ist die Studie von Augst und Faigel (1986) zur Entwicklung argumentativer Schreibkompetenzen. Doch wie lässt sich dieser hohe Stellenwert von Texten und Textkorpora erklären?

Dazu muss zunächst bedacht werden, dass die Schreibkompetenz eine *Eigenschaft von Individuen* ist (→ 2). Sie betrifft „Merkmale der schreibenden Person [...], die ihr innewohnen und die Ausgestaltung von Schreibprozessen steuern“ (Philipp, 2015, S. 9f.). Diese Merkmale können unterschiedlich modelliert und untersucht werden: „Schreibkompetenz ist genau das, als was sie theoretisch und methodisch bestimmt wird. Jede der beteiligten Wissenschaften macht sich davon ein eigenes Bild.“ (Feilke, 2015, S. 60). Dabei ist das Problem zu lösen, dass Kompetenzen nicht direkt beobachtbar sind und außerdem als relativ überdauernd angenommen werden. Deshalb müssen *manifeste Indikatoren* ausgewählt werden, die es erlauben, zu bestimmten Messzeitpunkten vergleichsweise zuverlässig auf die dahinter liegenden Kompetenzen zu schließen (Grabowski, 2014, S. 19).

Solche Indikatoren sind Prozessindikatoren (→ 20) oder Textindikatoren. Wie kompetent ein Schreiber oder eine Schreiberin ist, lässt sich am Schreibprozess, aber immer auch am Schreibprodukt erkennen. Dieses Produkt kann unterschiedliche Formen haben, es muss nicht zwingend ein Text sein (→ 9). Der Text aber ist gerade in Bildungskontexten nach wie vor das einschlägige Schreibprodukt – von der Primarstufe über die Sekundarstufen bis hin zum Studium (vgl. KMK, 2004; 2005; 2012). Ähnlich verhält es sich mit der Kommunikation. Das Schreiben und damit auch das Schreibprodukt kann unterschiedliche Funktionen haben. Die kommunikative Funktion aber ist, neben der kognitiven Funktion (vgl. Steinhoff, 2014), sicher die wichtigste Funktion des Schreibens in bildungsbezogenen und lebensweltlichen Kontexten. Wenn man nun beide Aspekte, den Text und die Kommunikation, zusammenführt, lässt sich sagen, dass das Schreiben typischerweise ein komplexes sprachliches Handeln der schreibenden Person für einen Leser oder eine Leserin mittels eines Textes ist. Da die schreibenden und die lesenden Individuen in der Regel zeitlich und räumlich getrennt sind, gelingt diese Kommunikation nur dann, wenn der Text sprachlich so ausgestattet ist, dass er aus sich heraus die ihm zugeordnete Funktion erfüllt. Der Text erschöpft sich folglich nicht darin, das Resultat der produzentenseitigen Denk-

prozesse und Schreibhandlungen zu sein, sondern ist immer auch ein für sich stehendes *Sprachwerk* (Bühler, 1934), das als solches von Schreibkompetenz zeugt und entsprechend untersucht werden kann. So erschließt sich, warum die Linguistik die Schreibkompetenz als *Textkompetenz* versteht: „Die Perspektive der Linguistik ist die Perspektive auf den *sprachlichen* Text. Gefragt wird nach den sprachlichen Komponenten, aus denen Texte hergestellt und aufgebaut werden, und nach der dafür geforderten Textkompetenz.“ (Feilke, 2015, S. 60).

Die Besonderheit der schreibdidaktischen Erforschung von Textkorpora besteht vor diesem Hintergrund darin, dass nicht einfach nur Texte untersucht werden, sondern Texte als *Indikatoren für die Schreibkompetenz eines Individuums*. Ob sich ein Textkorpus als empirische Grundlage einer schreibdidaktischen Studie eignet, ist folglich daran zu messen, inwieweit es den *authentischen Sprachgebrauch der avisierten Lernergruppe im interessierenden Schreibkompetenzbereich* abbildet. Textkorpora, die in der schreibdidaktischen Forschung untersucht werden, sind deshalb in aller Regel *Spezialkorpora*: Sie stehen nicht, wie *Referenzkorpora*, für eine Einzelsprache, sondern für bestimmte Sprachauschnitte (Lemnitzer & Zinsmeister, 2010, S. 106). Außerdem handelt es sich um *Lernerkorpora*: Sie repräsentieren spezifische Kompetenzausschnitte. Der Terminus „Lernerkorpus“ wird dabei in einem weiten, auf alle lernenden Personen bezogenen Sinne verstanden – anders als in der Korpuslinguistik, wo man ihn vor allem auf elektronische Korpora mit Texten von Zweit- und Fremdsprachenlernern bezieht (Granger, 2012).

Doch wie muss ein Textkorpus beschaffen sein und untersucht werden, damit dieser hohe Anspruch eingelöst wird? Diese Frage soll im Weiteren thematisiert werden, und zwar sowohl in Bezug auf Fälle, in denen mit einem eigenen Korpus gearbeitet wird, als auch im Hinblick auf Fälle, in denen fremde Textkorpora genutzt werden. Dazu wird zunächst auf die Korpusgestaltung, dann auf die Korpusaufbereitung und schließlich auf die Korpusauswertung eingegangen. Dieser Dreischritt folgt einem deduktiven Ansatz, der vor allem Strukturierungszwecken dient und nicht die Forschungspraxis abbilden soll, in der häufig eher induktiv vorgegangen wird. Das Ende des Beitrags bildet eine Zusammenfassung verschiedener Desiderate.

2. Gestaltung von Textkorpora

Wenn man mit einem Textkorpus arbeiten möchte, stellt sich zu Beginn die Frage nach seiner Gestaltung. Was macht ein gutes Korpusdesign aus? Worauf sollte man achten, wenn man ein eigenes Textkorpus zusammenstellt? Und was ist zu berücksichtigen, wenn man ein fremdes Textkorpus untersuchen will? Einige wichtige Punkte, die dabei zu beachten sind, werden im Folgenden erläutert. Im Einzelnen geht es darum, (2.1) welches Erkenntnisinteresse verfolgt wird und welche Ressourcen zur Verfügung stehen, (2.2) welche Kriterien bedeutsam sind,

(2.3) wie groß bzw. repräsentativ das Korpus sein sollte und (2.4) welche verfügbaren Korpora genutzt werden können.

2.1 Erkenntnisinteresse und Ressourcen

Die Korpusgestaltung muss zum einen zum *Erkenntnisinteresse* und den zugehörigen Forschungsfragen, Variablen und Konstrukten passen: „It is a truism that there is no such thing as a ‚good‘ or a ‚bad‘ corpus, because how a corpus is designed depends on what kind of corpus it is and how it is going to be used“ (Hunston, 2008, S. 155). Ein Textkorpus ist also dann angemessen gestaltet, wenn es seinen Zweck für die jeweilige Studie erfüllt. Will man mit einem eigenen Korpus arbeiten, muss das Design folglich auf das Erkenntnisinteresse abgestimmt sein. Möchte man hingegen ein fremdes, bereits vorliegendes Korpus untersuchen, muss das Erkenntnisinteresse an das Erkenntnispotenzial des Korpus angepasst werden.

Zum anderen hängt das Korpusdesign von den *Ressourcen* ab, über die man verfügt, z.B. den personellen, finanziellen und zeitlichen Möglichkeiten: „All corpora are a compromise between what is desirable [...] and what is possible“ (ebd., S. 156). Forschungsprojekte mit sehr umfangreichen Korpora werden deshalb häufig von größeren Forscherteams verantwortet (Grabowski, Becker-Mrotzek, Knopp, Jost & Weinzierl, 2014). Wenn nur wenige Ressourcen vorhanden sind, sollte das Erkenntnisinteresse entsprechend eingegrenzt und der Aufwand überschaubar gehalten werden, z.B. durch die Verwendung eines verfügbaren Korpus (vgl. Abschnitt 2.4).

2.2 Kriterien

Um zu beurteilen, ob ein Textkorpus zum Erkenntnisinteresse passt und umgekehrt, ist es sinnvoll, aus diesem Interesse bestimmte *Kriterien* abzuleiten und anzulegen. Zu diesem Zweck kann man sich am didaktischen Dreieck und mithin an drei Dimensionen orientieren – der *Sach-*, *Subjekt-* und *institutionellen Dimension* (Fix, 2008). Die Sachdimension betrifft den untersuchten Lerngegenstand, die Subjektdimension die erforschte Lerngruppe und die institutionelle Dimension den relevanten Lehr-Lern-Kontext.

2.2.1 Sachdimension

Die Sachdimension bezieht sich auf den *Lerngegenstand*, d.h. den untersuchten Sprach- und Kompetenzausschnitt. Dieser Ausschnitt muss sinnvoll konturiert sein, beispielsweise in Bezug auf Schreibfunktionen (z.B. Überzeugen), Textklassen (z.B. Anleitung), Schreibmedien (z.B. Wiki) oder Textprozeduren (z.B. Vergleichen). Dabei spielt auch die vertretene sprach- und kompetenztheoretische

Position eine Rolle. Es kann einen großen Unterschied für die Forschungsfrage und damit auch für die Korpusgestaltung machen, ob die Schreibkompetenz eher als generalisierte Kompetenz mit allgemeinen Merkmalen (z.B. Syntax) oder eher als schreib- bzw. textspezifische Kompetenz mit spezifischen sprachlichen Merkmalen (z.B. Textprozeduren) aufgefasst wird (Feilke, 2014).

In der Studie von Augst, Disselhoff, Henrich, Pohl und Völzing zur ontogenetischen Entwicklung der Text-Sorten-Kompetenz von der 2. bis zur 4. Jahrgangsstufe etwa wird der Lerngegenstand gleich im ersten Satz mit Bezug zu Textklassen und einem textspezifischen Kompetenzmodell expliziert (2007, S. 15): „Die vorliegende Studie untersucht die Entwicklung der produktiven, medial wie konzeptionell schriftlichen Text-Sorten-Kompetenz während des Grundschulalters und dies für die Textsorten: *Erzählung, Bericht, Instruktion, Beschreibung* und *Argumentation*“. Der so bereits klar umrissene Gegenstand wird im Verlauf der Arbeit mithilfe verschiedener Konstrukte operationalisiert. Um die Text-Sorten-Kompetenz der Schüler und Schülerinnen zu erfassen, werden schließlich verschiedene quantitative, grammatische, semantische und pragmatische Textmerkmale untersucht.

2.2.2 Subjektdimension

Die Subjektdimension bezieht sich auf die *Lernenden*, d.h. die personenbezogenen Merkmale, die für die Studie als Variablen relevant sind, z.B. das Alter, das Geschlecht, der familiäre Sprachgebrauch oder verschiedene kognitive, motivationale, soziale und ökonomische Voraussetzungen. Mit dem Alter etwa kann das biologische Alter der schreibenden Person, aber auch ihr kulturelles *Schreibalter* gemeint sein, d.h. „die praktische Schreiberfahrung und die Dauer der Auseinandersetzung mit den Standards und Normen einer literalen Kultur“ (Feilke, 1996, S. 1181; → 5). Die jeweiligen Daten können beispielsweise mithilfe von Fragebögen, Interviews oder Sprachtests erhoben werden.

In der Studie von Steinig, Betzel, Geider und Herbold (2009) zur diachronen Entwicklung von Schreibkompetenzen in der 4. Jahrgangsstufe (1972 und 2002) etwa werden mithilfe von Lehrer- und Elternfragebögen verschiedene personenbezogene Merkmale untersucht, von denen die Autoren annehmen, dass sie in einem systematischen Bezug zur Schreibkompetenz stehen: die Zweisprachigkeit, die soziale Schicht, die Schulempfehlung, das Geschlecht, das Alter und die Schulklasse. Eine spezifische Option, die sich im Zusammenhang der Altersvariable anbietet, besteht darin, ein Lernerkorpus mit einem *Expertenkorpus* zu vergleichen. So erhält man ein authentisches Maß zur Einschätzung der Qualität der Lernertexte. In Steinhoffs (2007) Studie zur Entwicklung wissenschaftlicher Schreibkompetenzen beispielsweise wird zu diesem Zweck ein Korpus studentischer Hausarbeiten mit einem Korpus von Fachaufsätzen professioneller wissenschaftlicher Autorinnen und Autoren verglichen.

Vor einer besonderen Herausforderung steht man, wenn die Erhebung personenbezogener Daten nicht oder nur bedingt möglich ist – z.B. bei der Untersuchung des Schreibens in der lebensweltlichen computervermittelten Kommunikation. In der Studie von Dürscheid, Wagner und Brommer (2010) zu Schreibkompetenzen in der Sekundar- und Berufsschule wird u.a. ein „Freizeitkorpus“ untersucht, das sich aus Texten zusammensetzt, die von den Schülern und Schülerinnen in ihrer Freizeit in neuen Medien geschrieben wurden (z.B. Chatbeiträge). Diese Texte liegen nur deshalb vor, weil sie freiwillig zur Verfügung gestellt wurden. Da nicht alle Schüler und Schülerinnen an der parallel durchgeführten Fragebogenerhebung teilgenommen haben und viele Texte nicht zweifelsfrei den Individuen zugeordnet werden konnten, ist für einige der kompetenzbezogenen Fragen der Studie nur ein Teil des Korpus nutzbar. Dies verdeutlicht exemplarisch die Schwierigkeiten, die sich in derartigen Fällen ergeben können – und die Notwendigkeit, sinnvolle Kompromisse zu finden.

2.2.3 Institutionelle Dimension

Die institutionelle Dimension bezieht sich auf den *Lehr-Lern-Kontext*, in dem die Korpus Texte produziert werden, z.B. auf die Bildungskultur, die Bildungseinrichtungen, die Lehrpersonen, die Curricula und den Unterricht. Dieser Kontext hat einen großen Einfluss auf die Schreibprozesse und Texte (Pohl & Steinhoff, 2010). Erhebt man ein eigenes Textkorpus, muss man deshalb genau darauf achten, den Lehr-Lern-Kontext so zu gestalten, dass die Lernenden ihr Leistungsvermögen tatsächlich abrufen können. Arbeitet man mit einem fremden Textkorpus, ist es wichtig, die Qualität dieses Kontextes differenziert einzuschätzen und die Texte vor dem Hintergrund dieser Einschätzung zu untersuchen.

In der Interventionsstudie von Anskeit (Anskeit & Steinhoff, 2014) zur Förderung deskriptiver und argumentativer Schreibkompetenzen in der 4. Klasse beispielsweise werden Schreibaufgaben eingesetzt, die zum Erkenntnisinteresse passen und zugleich sehr motivierend sind. Die erste Aufgabe besteht darin, ein Playmobil-Zimmer so zu beschreiben, dass es nachgebaut werden kann, die zweite Aufgabe darin, für eines von zwei möglichen Zielen für einen echten Klassenausflug zu argumentieren. Dabei findet eine Orientierung am pragmalinguistisch fundierten Konzept der *Aufgaben mit Profil* von Bachmann und Becker-Mrotzek (2010) und den dort genannten Merkmalen *Funktion, Interaktion, Wirkung, Weltwissen* und *Sprachwissen* statt (→ 10).

2.3 Größe und Repräsentativität

Ein Textkorpus ist auch durch seine *Größe* geprägt. In der Korpuslinguistik dominiert die Maxime „more data is better data“ (Church & Mercer, 1993, S. 18–19). Man bevorzugt Korpora mit mehreren Tausend Texten, weil man davon ausgeht,

dass sie differenziertere und besser verallgemeinerbare Erkenntnisse über den Sprachgebrauch ermöglichen. Das heißt natürlich keineswegs, dass nicht auch die Analyse kleinerer Korpora interessante Ergebnisse erbringen kann (Ghadessy, Henry & Roseberry, 2001).

In der schreibdidaktischen Forschung im deutschsprachigen Raum werden Textkorpora mit sehr unterschiedlichem Umfang erforscht. Konzentriert man sich auf Spezialkorpora, die für schreibdidaktische Projekte relevant sind, so reicht das Spektrum von unter zehn bis hin zu 20.000 Texten. Sehr große Korpora sind allerdings selten – auch in internationalen Studien (→ 11). Fünf der gegenwärtig bzw. in Kürze verfügbaren größten deutschsprachigen Lernerkorpora stammen aus der *DESI*-Studie (ca. 20.000 Texte: Neumann & Lehmann, 2008), dem *Schreibkompetenz*-Projekt (ca. 6.000 Texte: Grabowski et al., 2014), der *LAUII*-Studie (ca. 3.500 Texte: Neumann, 2007), dem *SimO*-Projekt (ca. 2.800 Texte; → 15) und der *Ludwigsburger Aufsatzkorpus*-Studie (ca. 2.300 Texte: Fix & Melenk, 2001). Ein Sonderfall ist das *Dortmunder Chat-Korpus* mit ca. 140.000 Beiträgen (Beißwenger, 2013).

Bei der Frage der *Repräsentativität* wird in der Korpuslinguistik inzwischen mehrheitlich davon ausgegangen, dass auch die umfangreichsten Korpora diese Bedingung im strikt definierten Sinne nicht erfüllen können (Perkuhn, Keibel & Kupietz, 2012, S. 46f.). Dies führt man darauf zurück, dass die dafür erforderliche, sowohl merkmalsgerechte als auch proportionsgetreue Bestimmung der Grundgesamtheit der Daten bei einer derart komplexen Materie wie einer Sprache bzw. eines Sprachausschnitts weder theoretisch noch praktisch erreicht werden kann. Um dennoch mit einer gewissen Legitimation Verallgemeinerungen vornehmen zu können, wird eine möglichst plausible *Balancierung* angestrebt (Kennedy, 1998, S. 62–66). Damit ist eine theoretisch stringente und ausgewogene Zusammenstellung relevanter Varianten einer Sprache bzw. eines Sprachausschnitts gemeint. Dabei werden *externe* und *interne* Kriterien angelegt (Miller & Biber, 2015, S. 33–35). Externe Kriterien betreffen die „situational representativeness“, also die Auswahl der Daten hinsichtlich der relevanten Domäne(n). Interne Kriterien beziehen sich demgegenüber auf die „linguistic representativeness“, also die Zusammenstellung der Daten bezüglich der interessierenden sprachlichen Formen und Funktionen.

Kritisch anzumerken ist, dass in einem Teil der deutschsprachigen Korpusstudien bisweilen ein etwas nachlässiger Umgang mit Fragen der Größe und Repräsentativität gepflegt wird. Ein typisches Argumentationsmuster besteht beispielsweise darin, dass zunächst auf bestimmte forschungspragmatische Restriktionen verwiesen wird, dann die angesichts dieser Restriktionen relative Vielzahl und Vielfalt der Korpus Texte betont wird, ohne die Auswahl hinreichend genau zu begründen, und schließlich die am Korpus gemachten Beobachtungen vergleichsweise unbefangenen verallgemeinert werden. Ein solches Vorgehen ist natürlich gerade bei Schlüssen von Texten auf Kompetenzen und Fördermaßnahmen einigermaßen heikel. Neuere Studien sollten sich deshalb um eine explizitere,

transparentere Erläuterung der Korpusbalancierung und eine nachvollziehbarere, vorsichtiger Generalisierung der Resultate bemühen.

2.4 Verfügbare Textkorpora

Es ist bereits thematisiert worden, dass eine sinnvolle Option für eine Korpusstudie, die mit begrenzten Ressourcen auskommen muss, in der Inanspruchnahme eines *verfügbaren Textkorpus* besteht. Von dieser Option wird in der Schreibdidaktik bislang nur selten Gebrauch gemacht (z.B. Gätje, Rezat & Steinhoff, 2012). Das ist durchaus verständlich, da (a) es nur wenige veröffentlichte Textkorpora gibt, (b) vielfach nur Korpusauszüge publiziert werden, (c) unklar ist, ob man die Daten nutzen darf und (d) die Texte oft suboptimal aufbereitet sind (vgl. Abschnitt 3). Daran lässt sich zugleich ablesen, dass die schreibdidaktische Forschung Korpora bislang primär aus Gründen der *Nachprüfbarkeit* zugänglich gemacht hat. Es wäre, gerade im Sinne des wissenschaftlichen Nachwuchses, wünschenswert, wenn sich eine ‚Kultur der Weiternutzung‘ etablieren würde, in der es selbstverständlich ist, dass Korpora veröffentlicht und dafür professionell aufbereitet werden. Ein positives Beispiel ist die Publikation des Korpus von Augst et al. (2007) auf einer eigens eingerichteten Webseite (s. Tab. 1).

Dass manche Korpora nicht oder nur in Auszügen veröffentlicht werden, hat bisweilen *rechtliche Gründe* (Lehberg, Chiarcos, Rehm & Witt, 2007). So dürfen in Deutschland Primärdaten, die durch das Urheberrecht geschützt sind, etwa literarische Texte, nur in geringem Umfang für einen kleinen forschenden oder studierenden Personenkreis publiziert werden. Die in der Schreibdidaktik untersuchten, selbst erhobenen Lernertexte hingegen können dann vollständig und für die Allgemeinheit veröffentlicht werden, wenn sie keine hinreichende *Schöpfungshöhe* haben, also einen geringen Individualitätsgrad aufweisen, und/oder wenn entsprechende Einverständniserklärungen der Lernenden bzw. ihrer Erziehungsberechtigten sowie der beteiligten Institutionen vorliegen.

Einige der bereits jetzt oder in Kürze verfügbaren deutschsprachigen Textkorpora mit einem gewissen Umfang, die sich für schreibdidaktische Zwecke anbieten, sind in Tabelle 1 aufgelistet. Um eine ungefähre Vorstellung davon zu vermitteln, ob sie für eine Weiternutzung in Frage kommen, werden Informationen zu relevanten Publikationen, Textklassen, Jahrgangsstufen bzw. Domänen sowie zur Textanzahl und gegebenenfalls zum digitalen Format gegeben. Da die Weiternutzungsfrage in den genannten Studien häufig nicht explizit thematisiert wird, empfiehlt es sich, vor einer Nutzung zunächst die Autoren bzw. Autorinnen und/oder den Verlag zu kontaktieren.

Bei einem der in der Liste angeführten Korpora handelt es sich um das derzeit bekannteste Lernerkorpus: das online nutzbare *Fehlerannotierte Lernerkorpus des Deutschen als Fremdsprache*, kurz *FaLKO* (Reznicek et al., 2012). Es setzt sich aus Texten von Studierenden mit Deutsch als Fremdsprache zusammen, die in verschiedenen Ländern erhoben wurden, und – zum Vergleich – aus Texten von Stu-

Tab. 1: Einige gegenwärtig und in Kürze verfügbare deutschsprachige Textkorpora.

Relevante Publikation	Textklasse(n)	Jahrgangsstufe(n) bzw. Domäne(n)	Textanzahl und ggf. digitales Format
Anskeit & Steinhoff, 2014	Argumentativer Brief, Raumbeschreibung	4. Jgst.	voraussichtlich ca. 1.000
Augst & Faigel, 1986	Argumentativer Brief	7., 10., 12. Jgst., Studierende	Anhang: 200 (inkl. Erweiterungskorpus)
Augst, Disselhoff, Henrich, Pohl, & Völzing, 2007	Argumentativer Brief, Bericht, Fantasieerzählung, Spielanleitung, Raum- beschreibung	2.–4., 6.–7., 9. Jgst.	Buchanhang: 90 (Auszug), online (.pdf): 1.210 (inkl. Erweiterungskorpus)
Ballis, 2010	Bericht, Beschreibung, Erzählung, Märchen, jeweils nach demselben Bildimpuls	5.–9. Jgst.	Beigelegte CD (.pdf): 333
Becker-Mrotzek, 1997	Bedienungsanleitung	4., 6., 7., 10., 12. Jgst., Studium, technische Redaktion	Buchanhang & online (.pdf): 39 (Auszug)
Beißwenger, 2013	Chat	Studium, andere Erwachsene	online (.html, .xml): 59.558 Beiträge
Bitter Bättig, 1999	Nacherzählung einer Bildergeschichte	4., 5., 6. Jgst.	Buchanhang: 24
Fix & Melenk, 2001	Fantasieerzählung nach Bildimpuls, Inhaltsangabe	8. Jgst.	Beigelegte CD (.rtf): 2.300
Grabowski et al., 2014	E-Mail, Kochrezept, Stellungnahme, Unfallbericht	5., 9. Jgst.	voraussichtlich ca. 6.000
Lehnen, 2000	Kurzartikel, Notiz, Textanalyse, Thesenpapier, Zusammenfassung	Studierende	Online (.pdf): 31
Knapp, 1997	Fantasieerzählung nach Bildimpuls	5.–7. Jgst.	Buchanhang: 56
Marx & Steinhoff (-> 15)	Figurenbeschreibung	6. Jgst.	voraussichtlich ca. 2.800
Neumann, 2007	Argumentativer Brief	9. Jgst.	Buchanhang: 36 (Auszug)
Petersen, 2014	Argumentativer Brief, Zusammenfassung	11., 13. Jgst., Studium	Buchanhang: 72
Rapti, 2005	Argumentativer Brief	4., 6., 8. Jgst.	Buchanhang: 126
Reznicek, Lüdeling, Krummes, Schwantuschke, Walter, Schmidt, Hirsch- mann & Andreas, 2012	Essay, Hausarbeit, Rezension, Zusammenfassung	Studium	Online (txt, .xml, .xlsx): ca. 700 (wird ausgebaut)
Schindler, 2004	Absagebrief, Bewerbungs- brief, Spielanleitung	Studierende	Online (.pdf): 44
Schmidlin, 1999	Nacherzählung einer Bildergeschichte	7.–11. Jgst.	Buchanhang: 124
Weinhold, 2000	Freier Text nach Bildimpuls	1. Jgst.	Buchanhang: 296
Winter, 1998	Bastelanleitung, Erlebniserzählung, freier Text nach Bildimpuls	4. Jgst.	Buchanhang: 302

dierenden mit einer deutschen Familiensprache. Die zentrale Besonderheit von FaLKo ist die Annotation von Lernerfehlern und „Zielhypothesen“ (vgl. Abschnitt 3.3). Hier wird an einem konkreten Beispiel deutlich, wie eine innovative schreibdidaktische Korpuspublikation aussehen könnte. Vielversprechend erscheint des Weiteren das (kurz vor der Veröffentlichung stehende) *KoKo-Projekt*, in dem ein L1-Korpus mit über 1.500 Erörterungen annotiert und analysiert wird, die von Schülern und Schülerinnen weiterführender Schulen in Thüringen und Tirol ein Jahr vor dem Abitur bzw. der Matura verfasst worden sind (Abel & Glaznieks, 2015).

Neben solchen Lernerkorpora können sich auch *Referenzkorpora* als nützlich erweisen. In Steinhoffs (2007) Studie beispielsweise wird das Korpus mit studentischen Hausarbeiten nicht nur, wie erwähnt, mit einem Korpus von Expertentexten verglichen, sondern auch mit einem Korpus journalistischer Texte, die aus dem *Deutschen Referenzkorpus (DeReKo)* des Instituts für Deutsche Sprache (IDS) stammen und online mithilfe des *COSMAS II*-Programms ausgewertet wurden. Ebenfalls als ertragreich könnten sich in Zukunft *Web-Korpora* herausstellen, d.h. Sammlungen von Internettexten, die mit bestimmten Kompilationsprogrammen generiert werden (Biber, Egbert & Davies, 2015).

3. Aufbereitung von Textkorpora

Wenn das Korpus gestaltet und erhoben worden ist, kann es auf verschiedene Weise aufbereitet werden, vor allem in Bezug auf (3.1) Primär- und Rohdaten, (3.2) Metadaten und (3.3) Annotationen. Diese Aufbereitung ist mit Blick auf das Erkenntnisinteresse und das Korpusdesign plausibel zu gestalten, mit Blick auf die Auswertung zweckmäßig anzulegen und mit Blick auf die Nachprüfbarkeit transparent zu dokumentieren.

3.1 Primär- und Rohdaten

Die wichtigsten Daten eines Korpus sind die *Primärdaten*, d.h. die Texte. Um die Primärdaten optimal auswerten zu können, wird häufig ein zweiter Datensatz mit sogenannten *Rohdaten* angelegt (Perkuhn et al., 2012, S. 54 ff.). Rohdaten sind vereinfachte, dekodierte Primärdaten, die nur diejenigen Informationen enthalten, die für das Erkenntnisinteresse relevant sind.

Ein für die Schreibdidaktik typisches Verfahren zur Erstellung von Rohdaten besteht darin, die zumeist handschriftlich vorliegenden Lernertexte abzutippen und dabei, je nach Erkenntnisinteresse, bestimmte Merkmale entweder bewusst auszusparen (z.B. Handschriften und Worttrennungen) oder genau zu dokumentieren (z.B. Überschriften und Abbildungen, Becker-Mrotzek, 1997, S. 204 ff.). Wenn es mit dem Erkenntnisinteresse vereinbar ist, können sogar deutliche Eingriffe in die Primärtexte vorgenommen werden, beispielsweise durch orthogra-

phische Normalisierungen (Augst et al., 2007). Dies kann sich auch aus methodischen Gründen als sinnvoll erweisen, etwa dann, wenn die Texte computergestützt ausgewertet werden sollen. Dafür bedarf es natürlich eines gewissen Maßes an sprachlicher Korrektheit (z.B. ein eindeutig klassifizierbarer Wortstamm). Sinnvoll kann zudem eine Segmentierung der Texte in kleinere Einheiten sein, z.B. durch eine „Tokenisierung“, also eine Zählung der Wortformen (Perkuhn et al., 2012, S. 30 ff.). Eine weitere wichtige Frage betrifft die Rohdatenformate. Im deutschsprachigen Raum werden zumeist die Formate *DOC*, *RTF* oder *PDF* genutzt (vgl. Tab. 1). Diese Formate können zwar leicht gelesen werden, sind für eine computergestützte Auswertung aber deutlich weniger geeignet als Formate wie *TXT* oder *XML*.

3.2 Metadaten

Korpora werden in aller Regel durch *Metadaten* angereichert. Diese Daten enthalten diejenigen kontextuellen Informationen, die benötigt werden, um die Primär- bzw. Rohdaten differenziert zu interpretieren (Lemnitzer & Zinsmeister, 2010, S. 46). Sie beziehen sich auf das Korpus insgesamt oder auf die Einzeltexte. Wenn sie digital erstellt werden, geschieht dies entweder *embedded*, in den Rohtextdateien (z.B. via XML), oder *standoff*, in separaten Dateien (Lehberg et al., 2007, S. 98). In der Korpuslinguistik ist es vor allem im ersten Fall üblich, mit standardisierten Metadatensätzen zu arbeiten (Perkuhn et al., 2012, S. 63 f.).

Einschlägige Metadaten für schreibdidaktische Korpusprojekte ergeben sich aus den in Abschnitt 2 beschriebenen Kriterien der Sach-, Subjekt- und institutionellen Dimension. Die Herausforderung besteht darin, eine ebenso passende wie umsetzbare Auswahl von Metadaten vorzunehmen, zu erheben und zu dokumentieren. Dabei kann man sich z.B. an Metadatensätzen orientieren, wie sie für das Ludwigsburger Aufsatzkorpus (Fix & Melenk, 2001, S. 79 ff.) oder das Korpus von Steinig et al. (2009) verwendet worden sind. Im Ludwigsburger Aufsatzkorpus etwa sind die Lernertexte in Tabellen eingebettet, die Metadaten zur Klasse, zum Schüler bzw. zur Schülerin, zum Geschlecht, zur Schulart, zur Muttersprache, zum Geburtsdatum, zum Textproduktionsdatum und zur Schreibaufgabe enthalten.

3.3 Annotationen

Korpora können zudem mit *Annotationen* angereichert werden (Lemnitzer & Zinsmeister, 2010, S. 60 ff.). Annotationen sind sprachwissenschaftliche Metainformationen zu den Primärdaten, die unterhalb der Textebene liegen. Sie dienen der linguistischen und spracherwerbstheoretischen Klassifizierung dieser Daten und können sich auf unterschiedliche sprachliche Gegenstandsbereiche beziehen. Die Annotation auf Wortebene nennt man *Tagging*, die Annotation auf Satzebene

Parsing. Wenn die Korpustexte digitalisiert vorliegen, können Annotationen auch automatisch erzeugt werden, beispielsweise im Hinblick auf Lemmata und Wortarten. Dafür eignen sich Programme wie der *RFTagger* oder der *TreeTagger*.

Ein gutes Beispiel für eines der wenigen annotierten und zugänglichen Textkorpora aus dem deutschsprachigen Raum ist das bereits erwähnte *Dortmunder Chat-Korpus* (Beißwenger, 2013). Die in diesem Projekt vorgenommenen Annotationen betreffen Typen von Chatbeiträgen, Stilelemente und die Dokumente selbst. Das Korpus kann online wie offline vollständig eingesehen und mit eigens zu diesem Zweck entwickelten Rechercheprogrammen untersucht werden. Bei den bereits erwähnten Projekten *FaLKO* und *KoKo* verhält es sich ähnlich. Digitale Annotationen, wie sie in solchen Projekten eingesetzt werden, machen es möglich, relevante Textstellen binnen Sekunden zu ermitteln und zu analysieren – und zwar zu jeder Zeit und von jedem Nutzer.

Auch das Annotieren selbst kann sich als wertvolles heuristisches Verfahren erweisen. In dieser Hinsicht ist vor allem die korpuslinguistische Diskussion zur sogenannten *Fehlerannotation* interessant (Lüdeling & Hirschmann, 2015). Dabei handelt es sich um eine Annotation von Lernertextäußerungen, die von Normen der Korrektheit oder Angemessenheit abweichen. Wenn beispielsweise eine Siebtklässlerin in einer Argumentation *Ich finde Hausaufgaben für sinnvoll* schreibt (Augst & Faigel, 1986, S. 213), tun sich gewisse Interpretationsspielräume auf, die Entscheidungen über die anzulegenden Kategorien und die sogenannten *Zielhypothesen* erforderlich machen. Zielhypothesen sind Annahmen zur „korrekten“ Äußerung, von der die Lerneräußerung (mutmaßlich) abweicht. Solche Hypothesen können zahlreich sein (z.B. *Ich finde Hausaufgaben sinnvoll* oder *Ich halte Hausaufgaben für sinnvoll*). Probleme wie diese stellen eine zentrale Herausforderung für die Lernertextanalyse im Speziellen und die schreibdidaktische Forschung im Allgemeinen dar. Sie treten natürlich auch bei einer herkömmlichen Textanalyse auf, werden durch das digitale Annotieren aber besonders virulent und verlangen dort zudem nach spezifischen Methoden.

4. Auswertung von Textkorpora

Die Auswertung von Korpustexten, d.h. die Analyse und Interpretation der Daten, erfolgt im Hinblick auf das spezifische Erkenntnisinteresse der jeweiligen Studie und kann demzufolge mit unterschiedlichen linguistischen Methoden vorgenommen werden. Da das umfangreiche Spektrum dieser Methoden im vorliegenden Beitrag nicht dargestellt werden kann (→ 12, 13, 14), liegt der Fokus auf einer Skizze der neueren korpuslinguistischen Instrumente. Diese Instrumente werden in der schreibdidaktischen Forschung bis heute kaum genutzt. Eine der wenigen Ausnahmen ist die Studie von Steinig et al. (2009), für die das bereits erwähnte *COSMAS II*-Programm in Anspruch genommen und eine Kooperation mit Korpuslinguisten vom IDS eingegangen wurde.

Die korpuslinguistische Methodologie ist durch zwei große Paradigmen geprägt. Diese Paradigmen werden z.T. in Opposition zueinander gesehen, z.T. aber auch als einander ergänzend verstanden (Hardie & McEnery, 2010). Das erste Paradigma ist das *corpus based* oder *corpus-as-method* genannte Verfahren. Hier dient das Korpus als Datenbestand zur Bestätigung oder Widerlegung deduktiv generierter Theorien. Das zweite Paradigma ist das *corpus driven* oder *corpus-as-theory* genannte Verfahren. Hier wird das Korpus als Datenbestand zur induktiven Gewinnung von Theorien verstanden: „observation leads to hypothesis leads to generalization leads to unification in theoretical statement“ (Tognini-Bonelli, 2001, S. 85). Beide Paradigmen haben sich in verschiedenen Forschungsbereichen als sehr erkenntnistiftend erwiesen (vgl. die Vielfalt der Beiträge in den Zeitschriften „Corpora“ und „International Journal of Corpus Linguistics“). Diese Entwicklung beruht wesentlich auf technologischen Innovationen. Dabei sind vor allem Rechercheprogramme von Bedeutung, die eine automatische Auswertung von Korpus-texten ermöglichen, z.B. *AntConc*, *MonoConc Pro* und *WordSmith Tools*. *AntConc* etwa ist ein häufig genutztes Programm, weil es die gängigen Analysemethoden enthält, kostenlos ist, auf allen populären Betriebssystemen läuft, lokal installiert wird, offline genutzt werden kann und eine benutzerfreundliche Oberfläche besitzt. Es eignet sich, um verschiedenste Auswertungsfragen nach Frequenzen, Assoziationsbeziehungen und sprachlichen Strukturen zu beantworten (Römer & Wulff, 2010).

Mit diesem Programm kann man sich z.B. die *Konkordanz* der jeweils interessierenden sprachlichen Formen anzeigen lassen, d.h. eine vertikale Liste dieser Formen, die mitsamt zuvor festgelegten, größeren oder kleineren Satz- und Textkontexten dargestellt wird. Dies lässt sich an einem Beispiel veranschaulichen. Abbildung 1 zeigt das Ergebnis einer Auswertung eines Subkorpus aus der Studie von Augst et al. (2007). Das Korpus umfasst 117 instruktive Texte (Spielanleitungen) aus der 2., 3. und 4. Klasse, die zu diesem Zweck in *TXT*-Dateien umgewan-



Abb. 1: Beispiel für eine Korpusauswertung mit *AntConc*.

delt worden sind. Das abgebildete Fenster visualisiert den Gebrauch der textinitiiierenden Prozedur *als erstes*. Mit ihr wird (auf konzeptionell mündliche Weise) die temporale Struktur der Instruktion eingeleitet (Knopp, Jost, Linnemann & Becker-Mrotzek, 2014, S. 121ff.). Das Ergebnis lädt zu verschiedenen Interpretationen und Anschlussuntersuchungen ein.

Ein Beispiel wie dieses illustriert, dass korpuslinguistische Programme eine rasche, exhaustive und nachprüfbare Datenrecherche ermöglichen und damit einen wichtigen Beitrag zur wissenschaftlichen Güte einer Korpusstudie leisten können.

5. Desiderate

Die schreibdidaktische Forschung im deutschsprachigen Raum, vor allem die Schreibentwicklungsforschung, blickt bei der Untersuchung von Textkorpora auf eine gewisse Tradition zurück, lässt diesbezüglich aber auch einige Desiderate erkennen. Hinsichtlich der *Korpusgestaltung* ist festzuhalten, dass es an größeren Korpora mangelt, die Daten sorgfältiger balanciert werden könnten und eine vorsichtiger Generalisierung der Ergebnisse angebracht wäre. In Bezug auf die *Korpusaufbereitung* und *Korpusauswertung* ist auffällig, dass die Daten nur selten und oft nur in Auszügen sowie nicht besonders nutzerfreundlich publiziert werden, weshalb eine „Kultur der Weiternutzung“ wünschenswert wäre. Hinzu kommt, dass die Optionen, die der Computer im Allgemeinen und die Korpus-technologie im Speziellen eröffnet, z.B. digitale Rohtextformate, bestimmte Rechercheprogramme oder automatische Annotationen, kaum genutzt werden. Insgesamt gesehen könnte die schreibdidaktische Forschung also sehr davon profitieren, wenn sie in Zukunft intensiv mit der Korpuslinguistik zusammenarbeiten würde.

Webseiten

Online-Lernerkorpora

Augst u.a., 2007	http://www.text-sorten-kompetenz.de
Becker-Mrotzek, 1997	http://www.verlag-gespraechsforschung.de/2004/pdf/schreiben.pdf
Dortmunder Chat-Korpus FaLKo	http://www.chatkorpus.tu-dortmund.de https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko
Lehnen, 2000	http://bieson.ub.uni-bielefeld.de/volltexte/2004/495/index.html
Schindler, 2004	http://pub.uni-bielefeld.de/publication/2305451

Rechercheprogramme

COSMAS II	http://www1.ids-mannheim.de/kl/projekte/korpora/
MonoConc Pro	http://www.athel.com/mono.html
WordSmith Tools	http://www.lexically.net/wordsmith/

Annotationsprogramme

RFTagger

<http://www.cis.uni-muenchen.de/~schmid/tools/RFTagger/>

TreeTagger

<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>**Literatur**

- Abel, A. & Glaznieks, A. (2015). Wo Sprachkompetenzforschung auf Varietätenlinguistik trifft: Empirische Befunde aus dem Varietäten-Lernerkorpus „KoKo“. In A. Lenz, T. Ahlers & M. Glauninger (Hrsg.), *Dimensionen des Deutschen in Österreich. Variation und Varietäten im sozialen Kontext* (S. 257–282). Frankfurt am Main: Peter Lang.
- Anskait, N. & Steinhoff, T. (2014). Schreibarrangements für die Primarstufe. Konzeption eines Promotionsprojekts und erste Ergebnisse zum Gebrauch von Schlüsselprozeduren. In T. Bachmann & H. Feilke (Hrsg.), *Werkzeuge des Schreibens. Beiträge zu einer Didaktik der Textprozeduren* (S. 129–155). Stuttgart: Fillibach bei Klett.
- Augst, G., Disselhoff, K., Henrich, A., Pohl, T. & Völzing, P.-L. (2007). *Text-Sorten-Kompetenz. Eine echte Longitudinalstudie zur Entwicklung der Textkompetenz im Grundschulalter*. Frankfurt am Main: Peter Lang.
- Augst, G. & Faigel, P. (1986). *Von der Reihung zur Gestaltung. Untersuchungen zur Ontogenese der schriftsprachlichen Fähigkeiten von 13–23 Jahren*. Frankfurt am Main: Peter Lang.
- Bachmann, T. & Becker-Mrotzek, M. (2010). Schreibaufgaben situieren und profilieren. In T. Pohl & T. Steinhoff (Hrsg.), *Textformen als Lernformen* (S. 191–210). Duisburg: Gilles & Francke. Verfügbar unter: http://www.koebes.uni-koeln.de/pohl_steinhoff.pdf [31.12.2015].
- Ballis, A. (2010). *Schriftsprachliche Förderung von Jugendlichen mit Migrationshintergrund. Eine Studie über Schüler(innen) der Sekundarstufe I unter besonderer Berücksichtigung der Jahrgangsstufen 5 bis 7*. Baltmannsweiler: Schneider Hohengehren.
- Becker-Mrotzek, M. (1997). *Schreibentwicklung und Textproduktion. Der Erwerb der Schreibfähigkeit am Beispiel der Bedienungsanleitung*. Opladen: Westdeutscher Verlag.
- Beißwenger, M. (2013). Das Dortmunder Chat-Korpus. *Zeitschrift für germanistische Linguistik*, 41 (1), 161–164.
- Biber, D., Egbert, J. & Davies, M. (2015). Exploring the composition of the searchable web: A corpus-based taxonomy of web registers. *Corpora*, 10 (1), 11–45.
- Bitter Bättig, F. (1999). *Die Entwicklung der schriftlichen Erzählfähigkeit vom 4. bis 6. Primarschuljahr*. Bern: Lang.
- Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Jena: Fischer.
- Church, K. W. & Mercer, R. L. (1993). Introduction to the Special Issue on Computational Linguistic Using Large Corpora. *Computational Linguistics*, 19 (1), 1–24.
- Dürscheid, C., Wagner, F. & Brommer, S. (2010). *Wie Jugendliche schreiben. Schreibkompetenz und neue Medien*. Berlin: de Gruyter.
- Feilke, H. (1996). Die Entwicklung der Schreibfähigkeiten. In H. Günther & O. Ludwig (Hrsg.), *Schrift und Schriftlichkeit. 2. Halbband* (S. 1178–1191). Berlin: de Gruyter.
- Feilke, H. (2014). Begriff und Bedingungen literaler Kompetenz. In H. Feilke & T. Pohl (Hrsg.), *Schriftlicher Sprachgebrauch – Texte verfassen* (S. 33–53). Baltmannsweiler: Schneider Hohengehren.
- Feilke, H. (2015). Text und Lernen – Perspektivenwechsel in der Schreibforschung. In S. Schmolzer-Eibinger & E. Thürmann (Hrsg.), *Schreiben als Medium des Lernens. Kompetenzentwicklung durch Schreiben* (S. 47–71). Münster: Waxmann.

- Fix, M. (2008). *Texte schreiben. Schreibprozesse im Deutschunterricht*. 2. Auflage. Paderborn: Schöningh.
- Fix, M. & Melenk, H. (2001). *Schreiben zu Texten – Schreiben zu Bildimpulsen. Das Ludwigsburger Aufsatzkorpus. Mit 2300 Schülertexten, Befragungsdaten und Bewertungen auf CD-Rom*. Baltmannsweiler: Schneider Hohengehren.
- Gätje, O., Rezat, S. & Steinhoff, T. (2012). Positionierung. Zur Entwicklung des Gebrauchs modalisierender Prozeduren in argumentativen Texten von Schülern und Studenten. In H. Feilke & K. Lehnen (Hrsg.), *Schreib- und Textroutinen. Theorie, Erwerb und didaktisch-mediale Modellierung* (S. 125–153). Frankfurt am Main: Peter Lang.
- Ghadessy, M., Henry, A. & Roseberry, R. L. (2001). *Small corpus studies in ELT. Studies in corpus linguistics*. Amsterdam: Benjamins.
- Grabowski, J. (2014). Kompetenz: ein bildungswissenschaftlicher Begriff. In J. Grabowski (Hrsg.), *Sinn und Unsinn von Kompetenzen. Fähigkeitskonzepte im Bereich von Sprache, Medien und Kultur* (S. 9–28). Opladen: Budrich.
- Grabowski, J., Becker-Mrotzek, M., Knopp, M., Jost, J., & Weinzierl, C. (2014). Different approaches to the assessment of text quality. An empirical report. In J. Engberg, C. Heine, & D. Knorr (Hrsg.), *Methods in writing process research* (S. 147–165). Frankfurt am Main: Peter Lang.
- Granger, S. (2012). Learner corpora. In C. A. Chapelle (Hrsg.), *The encyclopedia of applied linguistics* (S. 3235–3242). Oxford: Wiley-Blackwell.
- Hardie, A. & McEnery, T. (2010). On two traditions in corpus linguistics, and what they have in common. *International Journal of Corpus Linguistics* 15 (3), 384–394.
- Hunston, S. (2008). Collection strategies and design decisions. In Lüdeling, A. & Kytö, M. (Hrsg.), *Corpus linguistics. An international handbook. Volume 1* (S. 154–168). Berlin: de Gruyter.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. New York: Routledge.
- KMK (2004). *Bildungsstandards im Fach Deutsch für den Mittleren Schulabschluss. Beschluss vom 4.12.2003*. München: Wolters Kluwer.
- KMK (2005). *Bildungsstandards im Fach Deutsch für den Primarbereich. Beschluss vom 15.10.2004*. München: Wolters Kluwer.
- KMK (2012). *Bildungsstandards im Fach Deutsch für die Allgemeine Hochschulreife (Beschluss der Kultusministerkonferenz vom 18.10.2012)*. Verfügbar unter: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Deutsch-Abi.pdf [31.12.2015].
- Knapp, W. (1997). *Schriftliches Erzählen in der Zweitsprache*. Tübingen: Niemeyer.
- Knapp, M., Jost, J., Linnemann, M. & Becker-Mrotzek, M. (2014). Textprozeduren als Indikatoren von Schreibkompetenz – ein empirischer Zugriff. In T. Bachmann & H. Feilke (Hrsg.): *Werkzeuge des Schreibens. Beiträge zu einer Didaktik der Textprozeduren* (S. 111–128). Stuttgart: Fillibach bei Klett.
- Lehmer, T., Chiarcos, C., Rehm, G. & Witt, A. (2007). Rechtsfragen bei der Nutzung und Weitergabe linguistischer Daten. In G. Rehm, A. Witt & L. Lemnitzer (Hrsg.), *Datenstrukturen für linguistische Ressourcen und ihre Anwendungen – Data Structures for Linguistic Resources and Applications. Proceedings of the Biennial GLDV Conference 2007* (S. 93–102). Tübingen: Narr.
- Lehnen, K. (2000). *Kooperative Textproduktion. Zur gemeinsamen Herstellung wissenschaftlicher Texte im Vergleich von ungeübten, fortgeschrittenen und sehr geübten SchreiberInnen*. Bielefeld: Universität Bielefeld. Verfügbar unter: <http://bieson.uni-bielefeld.de/volltexte/2004/495/index.html> [31.12.2015].

- Lemnitzer, L. & Zinsmeister, H. (2010). *Korpuslinguistik. Eine Einführung* (2., durchges. und akt. Auflage). Tübingen: Narr.
- Lüdeling, A. & Hirschmann, H. (2015). Error annotation. In S. Granger, G. Gilquin & F. Meunier (Hrsg.), *The Cambridge handbook of learner corpus research*. Cambridge: Cambridge University Press.
- Miller, D. & Biber, D. (2015). Evaluating reliability in quantitative vocabulary studies: The influence of corpus design and composition. *International Journal of Corpus Linguistics*, 20 (1), 30–53.
- Neumann, A. (2007). *Briefe schreiben in Klasse 9 und 11. Beurteilungskriterien, Messungen, Textstrukturen und Schülerleistungen*. Münster: Waxmann.
- Neumann, A. & Lehmann, R. H. (2008). Schreiben Deutsch. In E. Klieme (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch* (S. 89–103). Weinheim: Beltz.
- Perkuhn, R., Keibel, H. & Kupietz, M. (2012). *Korpuslinguistik*. Paderborn: Fink.
- Petersen, I. (2014). *Schreibfähigkeit und Mehrsprachigkeit*. Berlin: de Gruyter.
- Philipp, M. (2015). *Schreibkompetenz. Komponenten, Sozialisation und Förderung*. Tübingen: Francke.
- Pohl, T. & Steinhoff, T. (2010). Textformen als Lernformen. In T. Pohl & T. Steinhoff (Hrsg.), *Textformen als Lernformen* (S. 5–26). Duisburg: Gilles & Francke. Verfügbar unter: http://www.koebes.uni-koeln.de/pohl_steinhoff.pdf [31.12.2015].
- Rapti, A. (2005). *Entwicklung der Textkompetenz griechischer, in Deutschland aufwachsender Kinder. Untersucht anhand von schriftlichen, argumentativen Texten in der Muttersprache Griechisch und der Zweitsprache Deutsch*. Frankfurt am Main: Peter Lang.
- Reznicek, M., Lüdeling, A., Krummes, C., Schwantuschke, F., Walter, M., Schmidt, K., Hirschmann, H. & Andreas, T. (2012). *Das Falko-Handbuch. Korpusaufbau und Annotationen. Version 2.01*. Verfügbar unter: https://www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko/Falko-Handbuch_Korpusaufbau%20und%20Annotationen_v2.01 [31.12.2015].
- Römer, U. & Wulff, S. (2010). Applying corpus methods to written academic texts: Explorations of MICUSP. *Journal of Writing Research*, 2 (2), 100–127.
- Schindler, K. (2004). *Adressatenorientierung beim Schreiben. Eine linguistische Untersuchung am Beispiel des Verfassens von Spielanleitungen, Bewerbungsbriefen und Absagebriefen*. Frankfurt am Main: Peter Lang.
- Schmidlin, R. (1999). *Wie Deutschschweizer Kinder schreiben und erzählen lernen. Textstruktur und Lexik von Kindertexten aus der Deutschschweiz und aus Deutschland*. Tübingen: Francke.
- Steinhoff, T. (2007). *Wissenschaftliche Textkompetenz. Sprachgebrauch und Schreibentwicklung in wissenschaftlichen Texten von Studenten und Experten*. Tübingen: Niemeyer.
- Steinhoff, T. (2014). Lernen durch Schreiben. In H. Feilke & T. Pohl (Hrsg.), *Schriftlicher Sprachgebrauch – Texte verfassen* (S. 316–330). Baltmannsweiler: Schneider Hohengehren.
- Steinig, W., Betzel, D., Geider, F. J. & Herbold, A. (2009). *Schreiben von Kindern im diachronen Vergleich. Texte von Viertklässlern aus den Jahren 1972 und 2002*. Münster: Waxmann.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: Benjamins.
- Weinhold, S. (2000). *Text als Herausforderung. Zur Textkompetenz am Schulanfang. Mit 296 Schülertexten aus Klasse I*. Freiburg i. Br.: Fillibach.
- Winter, C. (1998). *Traditioneller Aufsatzunterricht und kreatives Schreiben. Eine empirische Vergleichsstudie*. Augsburg: Wißner.