

- Sternberg, Robert J. (2002), The theory of successful intelligence and its implications for language aptitude testing. In: Robinson, Peter (Hrsg.) (2002a), 13-44.
- Walter, Catherine (2004), Transfer of reading comprehension skills to L2 is linked to mental representations of text and to L2 working memory. *Applied Linguistics* 25: 3, 315-339.
- Wesche, Marjorie (1981), Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In: Diller, Karl C. (Hrsg.) (1981), 119-154.
- Wesche, Marjorie; Edwards, Henry & Wells, Winston (1982), Foreign language aptitude and intelligence. *Applied Psycholinguistics* 3: 2, 127-140.
- Williams, John N. (1999), Memory, attention, and inductive learning. *Studies in Second Language Acquisition* 21, 1-48.
- Williams, John N. & Lovatt, Peter (2003), Phonological Memory and Rule Learning. *Language Learning* 53, 67-121.

Anglistische Korpuslinguistik und Fremdsprachenforschung: Entwicklungslinien und Perspektiven

Joybrato Mukherjee¹

The present paper provides a selected overview of the state of the art of applied English corpus linguistics and its influence on – and potential for – corpus-informed language pedagogy. Starting off from a general assessment of the impact that the corpus revolution has already had on English language teaching (ELT), the focus of the main part of this paper is on some typical examples of corpus use in three language-pedagogically relevant areas: (1) using corpora for ELT (e.g. producing learner dictionaries); (2) using corpora in the ELT classroom (e.g. in data-driven learning); (3) using learner corpora. The present paper also sketches out some important avenues for future research, e.g. in the area of learner-corpus analysis.

1. Einführung: die Korpusrevolution in der anglistischen Sprachwissenschaft und die Englischdidaktik²

Es besteht kein Zweifel daran, dass die Erstellung und Analyse großer computerisierter Korpora in der anglistischen Sprachwissenschaft die Beschreibung der englischen Sprache in den vergangenen Jahrzehnten revolutioniert hat (vgl. Rundell & Stock 1992). Dabei gilt es zu berücksichtigen, dass in der modernen Korpuslinguistik nicht jede Sammlung von Texten als ein Korpus gilt. Vielmehr definiert etwa Francis (1982: 7) das linguistische Korpus als "a collection of texts assumed to be representative of a given language, dialect, or other subset of a language". Das Konzept der Repräsentativität spielt in der Korpuslinguistik eine Schlüsselrolle, weil es darauf abzielt, das Korpus als eine statistisch zuverlässige Stichprobe aus dem Sprachgebrauch insgesamt (bzw. einer bestimmten Sprachvariante oder -varietät) zu konstruieren. Obgleich die Erstellung eines im statistischen Sinne repräsentativen Korpus schwierig umzusetzen ist und im Hinblick auf einige Aspekte kritisch hinterfragt werden kann (vgl. Rieger 1979, Biber 1993), ist das Bemühen um ein repräsentatives Korpusdesign von entscheidender Bedeutung, weil nur dann von den Ergebnissen einer Korpusanalyse

¹ Korrespondenzadresse: Joybrato Mukherjee, Justus Liebig Universität Gießen, Institut für Anglistik, Lehrstuhl für Englische Sprachwissenschaft, Otto-Behaghel-Str. 10B, 35394 Gießen, Email: Joybrato.Mukherjee@anglistik.uni-giessen.de

² Der vorliegende Aufsatz geht auf Vorträge bei der LearnTec-Konferenz in Karlsruhe 2005, bei der Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft (DGfS) in Bielefeld 2006 sowie in der Ringvorlesung *Educational Linguistics* an der Justus-Liebig-Universität Gießen 2007 zurück.

auf den Sprachgebrauch insgesamt geschlossen werden kann, so etwa im Hinblick auf die Häufigkeiten von Wörtern und syntaktischen Strukturen (vgl. Mukherjee 2004a).

Die Korpuslinguistik hat sich seit den frühen 1960er Jahren als eine neue Disziplin in der anglistischen Sprachwissenschaft herausgebildet, unter anderem angetrieben durch die Erstellung des *Brown University Corpus of Present-Day American English* (vgl. Francis & Kučera 1964). Das *Brown Corpus* stellt das erste von vornherein in maschinenlesbarer Form erstellte Korpus des Gegenwartsendlichen dar; es umfasst eine Million Wörter aus unterschiedlichsten schriftsprachlichen Genres des amerikanischen Englisch und kann aufgrund seines Designs als 'repräsentativ' für geschriebenes amerikanisches Englisch der frühen 1960er Jahre gelten. Das in den 1990er Jahren erstellte *British National Corpus* (BNC) umfasst bereits 100 Millionen Wörter, und zwar 90% aus einer Vielzahl schriftsprachlicher Genres und 10% aus sehr vielen verschiedenen mündlichen Kommunikationssituationen (vgl. Burnard 2000, 2007). Die immense Größe des BNC wird anschaulicher, wenn man berücksichtigt, dass das Vorlesen des gesamten Korpus bei einer normalen Sprechgeschwindigkeit und acht Stunden Lesetätigkeit am Tag verschiedenen Schätzungen zufolge fast zehn Jahre erfordern würde (vgl. Stubbs 1995: 49). Solche enormen Datenmengen können nur mit Hilfe moderner computertechnologischer Möglichkeiten gespeichert und automatisch analysiert werden.

In der Tat ist die Entwicklung der modernen Korpuslinguistik von Beginn an mit den Fortschritten in der Computertechnologie und der Software-Entwicklung verknüpft gewesen. Biber, Conrad & Reppen (1998) erklären anschaulich, dass nur mit Hilfe digitaler Speichermedien große maschinenlesbare Korpora erstellt und diese nur mit Hilfe computergestützter Methoden zuverlässig analysiert werden können (etwa im Unterschied zu manuellen Analysen):

Computers make it possible to identify and analyze complex patterns of language use, allowing the storage and analysis of a larger database of natural language than could be dealt with by hand. Furthermore, computers provide consistent, reliable analyses (Biber et al. 1998: 4).

Die computergestützte Analyse von Korpora, insbesondere mit Hilfe von Wortlisten, Konkordanzen, und anderen Instrumenten (etwa zur Identifikation von Kollokationen und *multi-word clusters*), die von modernen Software-Paketen wie z.B. *WordSmith Tools* (Scott 2005) angeboten werden, hat völlig neue Möglichkeiten der automatisierten empirischen Sprachbeschreibung ermöglicht, weil eine lineare manuelle Textanalyse nicht mehr nötig ist: "The corpus is stored in such a way that it can be analysed non-linearly, and both quantitatively and qualitatively" (Hunston 2002: 2).

Der vorliegende Aufsatz kann keinen vollständigen Überblick über die theoretisch-konzeptionellen Grundlagen, die Methoden und Techniken der Korpuslinguistik und die Vielfalt der korpusbasierten Erkenntnisse anbieten.³ Vielmehr sollen die wesentlichen Entwicklungslinien der Korpuslinguistik sowie zukünftige Perspektiven aus dem Blickwinkel der Fremdsprachenforschung vorgestellt und diskutiert werden. Da auch heute noch vor allem in der englischen Sprachwissenschaft korpuslinguistisch gearbeitet wird, ist es nahe liegend, dass der vorliegende Aufsatz sich auf die Anglistik bezieht – die linguistischen und fremdsprachendidaktischen Überlegungen sind allerdings, *mutatis mutandis*, auch auf das Lehren und Lernen anderer Sprachen übertragbar, wie neuere Diskussionen zu den Möglichkeiten und Grenzen der Korpuslinguistik in anderen Philologien wie der Romanistik (vgl. z.B. Pusch & Raible 2002) und Deutsch als Fremdsprache (vgl. z.B. Fandrych & Tschirner 2007, Walter & Grommes 2008) zeigen.

Man kann den Einfluss, den die Korpuslinguistik auf das Lehren und Lernen des Englischen als Fremdsprache in den letzten beiden Jahrzehnten ausgeübt hat, kaum überschätzen. Grundsätzlich lässt sich zwischen zwei Bereichen unterscheiden, in denen sich dieser Einfluss manifestiert: (a) der Inhalt des Englischunterrichts; (b) die Methodologie im Englischunterricht. Hunston (2002) grenzt die beiden Bereiche wie folgt voneinander ab:

The development of corpora has the potential for two major effects upon the professional life of the language teacher. Firstly, corpora lead to new descriptions of a language, so that the content of what the language teacher is teaching is perceived to change in radical ways [...]. Secondly, corpora themselves can be exploited to produce language teaching materials, and can form the basis for new approaches to syllabus design and to methodology (Hunston 2002: 137).

In den folgenden Abschnitten soll eine Bestandsaufnahme über den Einfluss der Korpusrevolution auf die Didaktik und Methodik des Englischunterrichts in den beiden genannten Bereichen erfolgen: (a) die korpusbasierte Beschreibung des Englischen und die entsprechende Veränderung der Lehr- und Lerninhalte im Englischunterricht (*content*, vgl. Abschnitt 2.1); (b) die Nutzung von Korpusdaten und korpuslinguistischen Methoden im Englischunterricht selbst (*methodology*, vgl. Abschnitt 3). Es sei allerdings bereits an dieser Stelle darauf hingewiesen, dass die Nutzung von Korpora im Unterricht nach wie vor vorrangig von Linguisten propagiert wird (vgl. z.B. Hunston 2002, Mukherjee 2002, Sinclair 2004, Braun, Kohn & Mukherjee 2006, O'Keefe, McCarthy & Carter

3 Hierzu sei auf entsprechende Einführungs-, Überblicks- und Nachschlagewerke hingewiesen (z.B. Kennedy 1998, Meyer 2002, Sampson & McCarthy 2004, Lemnitzer & Zinsmeister 2006, McEneaney, Xiao & Tono 2006, Teubert & Krishnamurthy 2007, Gries 2008, Mukherjee in Vorbereitung).

2007) – es besteht eine große Lücke zwischen der Vielzahl der von korpuslinguistischer Seite gemachten Vorschläge zu korpusbasierten Methoden des Fremdsprachenlernens einerseits und dem sehr geringen Ausmaß, in dem diese Vorschläge von Lehrpersonen tatsächlich im Englischunterricht in Deutschland umgesetzt werden andererseits (vgl. Mukherjee 2004b, Kreyer 2007). Granger (2004) stellt entsprechend fest:

The main fields of application of corpus data are materials and syllabus design and classroom methodology. In all three fields, there is very active work in progress, but, with the exception of ELT dictionaries, the number of concrete corpus-informed achievements is not proportional to the number of publications advocating the use of corpora to inform pedagogical practice (Granger 2004: 136).

Auch Tribble (2000: 31) bestätigt diesen Eindruck: "[n]ot many teachers seem to be using corpora in their classrooms". Es gibt möglicherweise einen Graben zwischen dem Enthusiasmus von Korpuslinguisten zu den fremdsprachendidaktischen Einsatzmöglichkeiten von Korpora und der Zurückhaltung sehr vieler Lehrpersonen im Hinblick auf den Einsatz korpusbasierter Methoden in ihrem eigenen Unterricht. Es erscheint daher sinnvoll und notwendig zu sein, sehr viel stärker als bisher die eigentlichen Lehrenden korpuslinguistisch aus-, fort- und weiterzubilden, wenn die Korpusrevolution tatsächlich auch im Klassenraum ankommen soll (vgl. Abschnitt 2.2). Im letzten Teil werde ich schließlich gesondert auf einen vergleichsweise jungen Zweig der Korpuslinguistik, nämlich die Erstellung und Analyse von Lernerkorpora, eingehen (vgl. Abschnitt 4).

2. Korpusbasierte Sprachbeschreibung und der Inhalt des Englischunterrichts

2.1 Wörterbücher, Grammatiken, Lehrwerke

Der früheste und wohl stärkste Einfluss der Korpuslinguistik auf die Fremdsprachenforschung lässt sich in der Lexikographie, d.h. in der Erstellung von korpusbasierten (Lerner-)Wörterbüchern, verorten. In den frühen 1980er Jahren begannen die Universität Birmingham und das Verlagshaus Collins in einem großen lexikographischen Projekt unter Leitung von John Sinclair, ein zunächst 20 Millionen Wörter umfassendes Korpus des Gegenwartsenglischen zu erstellen: die *Collins Birmingham University International Language Database* (COBUILD). Dieses Korpus bildete die Datengrundlage für das weltweit erste, komplett und ausschließlich korpusbasierte *COBUILD English Language Dicti-*

onary (1987), das in den Folgejahren neue Standards für die Wörterbuchschriftung insgesamt gesetzt hat: Heute ist nahezu jedes englischsprachige Wörterbuch aller großen Verlagshäuser korpusbasiert. Das COBUILD-Korpus wurde seit den 1980er Jahren kontinuierlich ausgebaut und umfasst heute – unter der Bezeichnung *Bank of English* – bereits mehr als eine halbe Milliarde Wörter.

Korpusbasierte Wörterbücher enthalten in vielerlei Hinsicht Informationen über den Sprachgebrauch, wie sie in traditionellen Wörterbüchern aus der Vor-Korpus-Zeit nicht vorhanden waren. Der folgende Eintrag aus dem korpusbasierten *Macmillan English Dictionary* (2002) für das Substantiv *assumption* veranschaulicht beispielhaft die lexikographischen Implikationen der Korpusrevolution:

assumption /ə'sʌmpʃ(ə)n/ noun ★★

1 [C] something you consider likely to be true even though no one has told you directly or even though you have no proof: *Your argument is based on a completely false assumption.* ♦ **+that** *There is an assumption that all the people who live around here are rich.* ♦ **make an assumption** (=make a decision based on poor evidence) *People tend to make assumptions about you based on your appearance.* ♦ **on the assumption that** *The law works on the assumption that it is preferable for children to be with their mother.*
2 [U] a process in which you begin to use your power or status, or begin to be responsible for something: *the assumption of adult responsibilities* (*Macmillan English Dictionary* 2002, s.v. *assumption*)

Der Eintrag für *assumption* enthält Informationen über die Häufigkeit des Lexems im Englischen allgemein (über die Anzahl der Sternchen in der ersten Zeile), die relative Häufigkeit der einzelnen Bedeutungen von *assumption* (so ist die Bedeutung **1** häufiger als die Bedeutung **2**), die häufig mit *assumption* gemeinsam auftretenden Wörter (z.B. Kollokationen wie *assumption + that*) und die typischen lexikogramatischen *patterns* des Wortes (z.B. *on the assumption that*). Die Angabe von Kollokationen und lexikogramatischen *patterns*, in denen *assumption* signifikant häufig vorkommt, verdeutlicht im Übrigen das generelle Bemühen in der Korpuslinguistik, die traditionelle Grenze zwischen Lexik und Grammatik zu überwinden und statt dessen eine holistische lexikogramatische Beschreibung des Sprachgebrauchs anzustreben (vgl. Stubbs 1993). Neben Kollokationen und *patterns* enthalten korpusbasierte Wörterbücher oft auch Angaben zu anderen häufig wiederkehrenden Routinen im Gebrauch von Wörtern, z.B. so genannte *colligations* (d.h. das gemeinsame Auftreten von grammatischen Kategorien wie z.B. Wortklassen, so etwa die sehr häufige *colligation* 'Verb/Adjektiv + Präposition + bestimmter Artikel + *naked eye*', vgl. Sinclair 1991) und semantische Prosodien (d.h. die Tendenz eines Wortes, in positiven oder negativen Kontexten gebraucht zu werden, so z.B. das

Verb *provide* in positiven und das Verb *affect* in negativen Kontexten, vgl. Stubbs 1995).

Während korpusbasierte Wörterbücher zunehmend grammatische Angaben zur Verwendung einzelner Wörter beinhalten, ergänzen korpusbasierte Grammatiken im Sinne des lexikogrammatischen Ansatzes die Beschreibung syntaktischer Strukturen um relevante lexikalische Informationen, so etwa Wortlisten mit denjenigen Wörtern, die typischerweise in einer bestimmten syntaktischen Struktur vorkommen. Die erste korpusbasierte Grammatik des Englischen ist ebenfalls aus dem COBUILD-Projekt hervorgegangen. Wie die COBUILD-Wörterbücher ist auch die *Collins COBUILD English Grammar* (1990) als ein Referenz- und Nachschlagewerk für Sprachenlerner konzipiert worden. Entsprechend ist die Beschreibung der grammatischen Kategorien, die in der COBUILD-Grammatik angesetzt werden, ausgesprochen einfach verständlich, nicht terminologisch überladen und prinzipiell semantisch orientiert. Der folgende Eintrag zu so genannten *emphasizing adjectives* zeigt beispielhaft, dass auch die Lernergrammatikschreibung sehr von der Korpuslinguistik profitiert hat, da jetzt Lerner des Englischen syntaktische Strukturen direkt gemeinsam mit den signifikant häufig in ihnen auftretenden lexikalischen Elementen lernen können:

You can emphasize your feelings about something that you mention by putting an adjective such as 'complete', 'absolute', and 'utter' in front of a noun.

He made me feel like a complete idiot.

Some of it was absolute rubbish.

...utter despair.

...pure bliss.

You generally use an adjective of this kind only when the noun indicates your opinion about something. Because they are used to show strong feelings, these adjectives are called **emphasizing adjectives**. Here is a list of emphasizing adjectives:

absolute	outright	pure	true
complete	perfect	real	utter
entire	positive	total	

(*Collins COBUILD English Grammar* 1990: 69)

In den 1990er Jahren wurden weitere korpusbasierte Grammatiken erstellt. Das wichtigste und umfangreichste Beispiel für eine korpusbasierte Grammatik ist die *Longman Grammar of Spoken and Written English* (Biber, Johansson, Leech, Conrad & Finegan 1999), eine ausgesprochen detaillierte Referenzgrammatik mit einer Fülle an quantitativen Informationen zur Häufigkeit und Verteilung syntaktischer Strukturen in vier großen Textsorten des Gegenwarts-englischen: Konversation, Zeitungssprache, Wissenschaftssprache und die Sprache der Literatur. Daneben sei ebenso auf *An Empirical Grammar of the English Verb System* (Mindt 2000), eine korpusbasierte Beschreibung der englischen

Verbalphrase, hingewiesen wie auch auf die *online* über das Internet verfügbare *Chemnitz Internet Grammar of English* (<<http://www.tu-chemnitz.de/InternetGrammar/>>, vgl. Schmied 1999, Schmied & Haase 2003), in der auf der Grundlage eines deutsch-englischen Übersetzungskorpus aus kontrastiv-linguistischer Sicht wichtige Fehlerquellen für deutsche Lerner des Englischen behandelt werden. Es soll auch nicht unerwähnt bleiben, dass korpuslinguistische Erkenntnisse inzwischen auch Einfluss auf die Gestaltung normaler Lernergrammatiken deutscher Verlage ausüben. So verweist Ungerer (1999) im Vorwort zu *Englische Grammatik heute* darauf, dass die Grammatik auch Erkenntnisse aus Analysen des BNC berücksichtigt habe.

Es muss in diesem Zusammenhang darauf hingewiesen werden, dass die aus der Analyse von sehr großen, repräsentativen Korpora gewonnenen Einsichten über den authentischen und typischen Sprachgebrauch – etwa auf der lexikalischen und der syntaktischen Ebene – häufig im Widerspruch zu dem stehen, was individuelle Sprachverwender, Lehrer und Linguisten intuitiv annehmen. Sehr viele Sprecher des Englischen im allgemeinen und Englischlehrer im Besonderen gehen zum Beispiel davon aus, dass in *if-clauses* (d.h. in Konditionalsätzen, die durch die Konjunktion *if* eingeleitet werden) das finite Verb nicht im *conditional* verwendet werden könne. Die entsprechende Schulregel "Kein *would* in *if-clauses*!" geht auf diese intuitionsbasierte Annahme zurück. Wenn man sich allerdings große Korpora des Gegenwartsenglischen ansieht, stellt man fest, dass diese Wahrnehmung nicht ganz richtig ist bzw. einer wichtigen Ergänzung bedarf: nach der Phrase *X would be grateful if* ist der Gebrauch von *would* im *if-clause* nicht nur möglich, sondern auch typisch und idiomatisch (vgl. Clear 2000). Die folgenden Beispiele stammen allesamt aus der britischen Komponente des *International Corpus of English* (ICE-GB), das 1 Million Wörter aus gesprochenen (60%) und geschriebenen (40%) Texten der frühen 1990er Jahre enthält (vgl. Nelson, Wallis & Aarts 2002):

- I would be very **grateful** if you would contribute in some way to the maintaining of the Squadron 's morale.
- I would be **grateful** if you would let me know how much it would cost to employ a First Aider on a daily basis
- I would be **grateful** if you would supply me with Soluprick wet skin prick test as listed on the enclosed order form for use on the above named patient.
- I enclose the correct record form, and would be **grateful** if you would complete it and send it back.
- I would be **grateful** if you would let me have smallpox vaccine for the abovenamed patient who is working with vaccinia virus at University College London and wishes to be vaccinated against smallpox.

Korpuslinguistische Befunde zeigen zum einen, dass individuelle Intuitionen und Regelvorstellungen der sprachlichen Realität widersprechen. Zum anderen ermöglichen sie auch einen empirischen Zugriff auf Aspekte des Sprachgebrauchs, die sich ganz grundsätzlich der Intuition entziehen, insbesondere Häufigkeiten von sprachlichen Formen und Strukturen. Es ist beispielsweise schlicht unmöglich, rein intuitiv die zehn häufigsten Wörter im Gegenwartsgenesischen oder die häufigste Bedeutung eines einzelnen Wortes zu benennen.

Wenn der typische Sprachgebrauch in einer Sprechergemeinschaft quantitativ-empirisch beschrieben werden soll, ist die Observation großer Mengen authentischer und natürlicher Sprache, wie sie über Korpora zugänglich gemacht werden, unumgänglich.

Während heute nahezu alle Lernerwörterbücher und zunehmend viele Lernergrammatiken korpuslinguistische Erkenntnisse berücksichtigen, wird die Sprache in Englischlehrbüchern für den Schulunterricht noch immer in aller Regel nicht mit dem tatsächlichen Sprachgebrauch abgeglichen. Englischlehrbücher müssen in vielerlei Hinsicht weiterentwickelt werden, damit die in ihnen enthaltene Sprache einen realistischen Eindruck von natürlichem und muttersprachlichem Englisch vermitteln kann. Als ein Beispiel für die Realitätsferne und Unnatürlichkeit der englischen Sprache in Englischlehrbüchern sei der Gebrauch von so genannten *discourse markers* genannt. *Discourse markers* sind Partikeln, die üblicherweise in spontan gesprochener Sprache verwendet werden, nicht Teil der syntaktischen Struktur sind und verschiedene pragmatische Funktionen im gesprochenen Diskurs erfüllen, z.B. *well, so, you know* und *like*. Müller (2005) zeigt im Zusammenhang ihrer korpusbasierten Untersuchung fortgeschrittener englischer Lernersprache, dass in verschiedenen Englischlehrwerken, die an vielen deutschen Schulen verwendet werden (nämlich *Learning English Green Line, English G and Notting Hill Gate*), *well* und *so* recht häufig als *discourse markers* vorkommen (bis zu 24 bzw. 20 Mal in einem Band), während *you know* und *like* als *discourse markers* nur ausgesprochen selten belegt sind (bis zu 6 bzw. 3 Mal in einem Band). Müller (2005) kann auf dieser Grundlage eine Erklärung dafür anbieten, warum die von ihr untersuchten fortgeschrittenen Englischlerner im *Giessen Long Beach Chaplin Corpus (GLBCC)* *you know* und *like*, die von Muttersprachlern sehr häufig verwendet werden, kaum einsetzen:

Given this representation of the four discourse markers in German textbooks of English, it is not surprising that the German speakers in the GLBCC did not have much difficulty using *well*, but apparently were not used to employing *you know* and *like* as much as native speakers did. It also fits the picture that *so* is less frequent in the German speaker data than in the native speaker data, since it seems to be underrepresented in the textbooks [...] (Müller 2005: 250).

Es kann in jüngster Zeit aber festgestellt werden, dass auch Lehrwerksentwickler zunehmend Korpusdaten berücksichtigen – hier sei insbesondere das neue Lehrwerk *Touchstone*, in das Informationen aus dem *Cambridge International Corpus* eingeflossen sind, genannt.

2.2 Aus- und Fortbildung von Englischlehrern

Die bislang nur geringe Berücksichtigung korpuslinguistischer Befunde in der Konzeption von Lehrwerken hat sicherlich auch damit zu tun, dass bis in die 1990er Jahre hinein die Korpuslinguistik im Grunde noch keinen systematischen und umfassenden Eingang in die Lehrpläne der anglistischen Studiengänge an deutschen Hochschulen gefunden hatte bzw. als Spezialisierungsgebiet ohne Weiteres durch andere sprachwissenschaftliche Teildisziplinen ersetzt werden konnte. Es ist daher nicht überraschend, dass noch vor fünf Jahren in einer Umfrage unter knapp 250 Englischlehrern an nordrhein-westfälischen Gymnasien etwa 80% der Befragten angaben, von linguistischen Korpora und korpuslinguistischen Methoden bislang noch nichts erfahren zu haben (vgl. Mukherjee 2004b: 241). Dieses Bild scheint sich in jüngster Zeit allerdings deutlich zu verändern, da seit den 1990er Jahren die computergestützte Korpuslinguistik zunehmend Teil der *mainstream linguistics* und daher auch zu einem integralen Bestandteil der anglistischen Hochschulausbildung geworden ist (auch im Zuge der Neuentwicklung von Studiengängen im Rahmen des Bologna-Prozesses) und weil in allen Bundesländern vermehrt jüngere Lehrkräfte eingestellt werden, die in ihrem Studium entsprechend Erfahrungen im Umgang mit Korpora und korpuslinguistischen Methoden haben sammeln können.

Es ist in diesem Zusammenhang ausgesprochen wichtig, Lehramtsstudenten von vornherein die zahlreichen Vorteile einer *corpus literacy* (Mukherjee 2002: 179f.) bzw. einer *corpus competence* (Kreyer 2008: im Druck) zu vermitteln. *Corpus literacy* meint zunächst die Fähigkeit, zur Klärung sprachlicher Fragen und zur Lösung linguistischer Probleme selbstständig und kompetent geeignete Korpora und relevante korpuslinguistische Programme nutzen zu können. Eine solche *corpus literacy* ist eine Grundvoraussetzung dafür, dass zukünftige Lehrpersonen Korpusdaten als eine wichtige Grundlage etwa zur Bewertung von sprachlicher Idiomatizität im lernersprachlichen *output* nutzen können. Deutsche Lerner des Englischen neigen beispielsweise dazu, aufgrund des im Deutschen inzwischen üblichen 'ein *statement* abgeben' auch im Englischen *give a statement* zu verwenden. Ist diese Verb-Nomen-Kollokation im Englischen möglich bzw. kommt sie vor? Ist sie, falls sie denn vorkommt, häufiger – und damit idiomatischer – als die Phrase *make a statement*? Zur Beantwortung dieser Fragen

genügt ein Blick in das BNC. Man stellt bereits bei einer einfachen Phrasensuche fest, dass *give/gives/giving/gave/given a statement* insgesamt nur 12 Mal in 100 Millionen Wörtern vorkommt, während *make/makes/making/made a statement* insgesamt 493 Mal belegt ist. Ohne auch noch weitere relevante Vorkommen der beiden Verb-Nomen-Kollokationen suchen zu müssen (z.B. Passivkonstruktionen und Vorkommen mit einem prämodifizierenden Adjektiv vor *statement*), belegt der eindeutige Frequenzunterschied, dass die idiomatischere Variante *make a statement* ist und entsprechend im Englischunterricht vermittelt werden sollte.

Die Korpusdaten zeigen aber auch, dass *give a statement* zwar vergleichsweise selten vorkommt, aber durchaus mehrfach belegt ist – und damit nicht ohne weiteres als ein Fehler markiert und bewertet werden kann. An diesem Beispiel zeigt sich eine grundsätzliche Herausforderung der Nutzung von Korpusdaten für die Fremdsprachendidaktik. Oftmals ist es nicht möglich, qualitativ-kategorisch zwischen "richtigem Sprachgebrauch" und "falschem Sprachgebrauch" zu unterscheiden – vielmehr sind es oft quantitativ-graduelle Unterschiede zwischen idiomatische(re)n Varianten und weniger idiomatischen Varianten, die in Korpora identifiziert werden können. Dretzke & Nester (2001) sprechen daher zu Recht davon, dass viele Bereiche in der Grammatik sich einer klaren Unterscheidung zwischen "richtig" und "falsch" entziehen und man das grammatische System eher als eine *fuzzy grammar* ansehen sollte. Daraus kann abgeleitet werden, dass das Korrekturverhalten von Lehrpersonen idealerweise in zwei Dimensionen ausdifferenziert werden sollte:

- Korrekturbreite → damit ist die Bewertung von lernersprachlichem *output* auf drei Ebenen der Adäquatheit des Sprachgebrauchs gemeint:
 - (a) lexikogrammatikalische Korrektheit (= in der Zielsprache mögliche Form/Struktur),
 - (b) Akzeptabilität (= Kontextangemessenheit),
 - (c) Idiomatizität (= muttersprachliche Präferenz);
- Korrekturtiefe → damit ist die Bewertung von lernersprachlichem *output* (auf jeder der o.g. drei Ebenen) im Hinblick auf die folgenden möglichen Beschreibungen gemeint:
 - (a) der *output* entspricht den Regeln und der Norm;
 - (b) der *output* entspricht nicht der Norm, stellt aber keinen systematischen Regelverstoß dar;
 - (c) der *output* entspricht weder den Regeln noch der Norm.

An anderer Stelle zeige ich am Beispiel des im informellen, mündlichen Sprachgebrauch vorkommenden *get-passive* (z.B. *it got lost* anstelle des standardsprachlichen *be-passive* in *it was lost*), dass in lernersprachlichem *output* grundsätzlich alle Kombinationen der Adäquatheit (im Sinne der Korrekturbreite) und der Grade der Regel-/Normentsprechung (im Sinne der Korrekturtiefe) angetroffen werden können (vgl. Mukherjee 2002: 140ff.). Eine nach den genannten Kategorien in den beiden Korrekturdimensionen differenzierte Lernersprachen- und Fehlerbewertung ist ohne Zweifel nur im fortgeschrittenen Fremdsprachenunterricht sinnvoll und angemessen. Insgesamt muss die Fremdsprachendidaktik sich im Lichte von Korpusdaten stärker der Frage widmen, wie sich die Fehlerbewertung und die dabei notwendigen kategorialen Unterscheidungen zwischen korrektem und fehlerhaftem Sprachgebrauch in Einklang mit der sehr viel differenzierteren Realität des Sprachgebrauchs in Korpora bringen lassen und welche Konsequenzen hieraus auch für die Aus- und Fortbildung von Lehrpersonen zu ziehen sind.

Corpus literacy umfasst über analytische und methodische Kompetenzen der Nutzung von Korpora und korpuslinguistischen Programmen auch die Fähigkeit, *ad hoc* eigene Korpora – so genannte "DIY ('do-it-yourself') corpora" (McEnery et al. 2006: 71) zu erstellen. Da das Internet eine Fülle von bereits computerisierten Daten anbietet, kann es für Lehrpersonen ausgesprochen nützlich sein, sich aus Internetdaten – soweit es Copyright-Restriktionen zulassen – für den Unterricht so genannte *web-derived corpora* zusammenzustellen, so z.B. Korpora mit *pop song lyrics* (vgl. z.B. Kreyer & Mukherjee 2007). In den Abschnitten 3 und 4 werden verschiedene DIY-Korpora vorgestellt, u.a. ein Lernerkorpus, das von einer Lehrperson selbst erstellt wurde, und zwar aus Texten, die von den Lernern in ihrem eigenen Englischkurs stammen. Auch ein solches "lokales" Lernerkorpus kann fremdsprachendidaktisch sinnvoll genutzt werden.

Wie verschiedene Studien zeigen, ist es aus Sicht der Lehrerausbildung und -fortbildung wichtig, die Vermittlung einer *corpus literacy* im Studium zukünftiger Lehrerinnen und Lehrer um korpuslinguistische Weiterbildungs- und Beratungsangebote für Lehrerinnen und Lehrer zu ergänzen, die sich bereits im Schuldienst befinden. In diesem Kontext sei ein sehr interessantes Projekt erwähnt, das in Hong Kong durchgeführt wurde: das *TeleNex Project* (vgl. Tsui 1996, 2004). Im Rahmen dieses Projekts wurde eine *website* für Englischlehrerinnen und -lehrer in Hong Kong entwickelt (unter <http://www.telenex.hku.hk>), die verschiedene Diskussionsforen enthält, u.a. eine so genannte *language corner*, in der Lehrpersonen Fragen zum korrekten Sprachgebrauch und Unsicherheiten bei der Bewertung von Lernersprache diskutieren und von korpuslinguistischen Experten (die unter anderem die *Bank of English* und das BNC zu Rate ziehen) klären lassen können. Die folgenden bei-

den Fragen von Lehrpersonen beziehen sich beispielsweise auf das Problem, ob mit koordinierten Nominalphrasen in der Subjektposition bzw. in der Konstruktion mit *existential there* die Singular- oder Pluralform des Verbs zu verwenden ist:

Teacher 4:
Hello! Which one is correct?
There is a man and a woman outside.
Or
There are a man and a woman outside.
Please give some comments, any one.

Teacher 5:
Hi,
What should we use in the following sentences? Is or are?
1. There _____ an apple and some oranges on the table.
2. There _____ some oranges and an apple on the table.
Thanks. It seems to me that 'are' is okay in both. Is there any rule here?

(Tsui 2004: 50)

Die Fragen belegen anschaulich, wie nützlich das Angebot der *language corner* ist, da Englischlehrerinnen und -lehrer, die selbst nicht Muttersprachler des Englischen sind, durchaus unsicher darüber sein können, wie der korrekte, akzeptable und/oder idiomatische Sprachgebrauch im Englischen ist.⁴ Die Lehrpersonen erhielten die folgende Antwort von den korpuslinguistischen Mitarbeitern des Projekts:

TELEC staff responded to the teachers' questions by pointing out that usually the singular form of 'be' is used when the first noun that follows is singular and the plural form of 'be' is used when the noun group after it is plural (see also *Collins Cobuild English Grammar*, p. 416). However, a search through the corpus does show an instance of the following:

According to PACE, suspects can only be detained at designated police stations where **there are a custody and a reviewing officer** (Tsui 2004: 51).

Einerseits kann man folglich durchaus eine Regel wie in der COBUILD-Grammatik formulieren, andererseits zeigt sich, dass in der Konstruktion mit *existential there* manchmal auch die Pluralform des Verbs *be* mit einem nachfolgenden Nomen im Singular verwendet wird. Das Beispiel verdeutlicht, dass grammatische Regeln zwar eine mehr oder weniger große Mehrheit aller Fälle

4 Die Thematisierung solcher sprachlicher Unsicherheiten auf Seiten der Lehrperson scheinen grundsätzlich – auch in Deutschland – ein Tabu darzustellen: Wenn es einzig die Lehrperson ist, die bei der Bewertung von Lernersprache über die Korrektheit bzw. die Wertung als Fehler entscheidet, ist es sicherlich schwierig, die fremdsprachliche Autorität (und Kompetenz) der Lehrperson zu hinterfragen.

erklären können, dass aber ein Rest verbleibt, der von den Regelbeschreibungen nicht abgedeckt wird, dabei aber nicht falsch ist:⁵ "the question is not about possibility but about probability of usage" (Tsui 2004: 51).

Während bisher die Nutzung von Korpora durch bzw. für Lehrpersonen außerhalb des eigentlichen Unterrichts im Vordergrund stand, wende ich mich im Folgenden der Nutzung von Korpora durch Schülerinnen und Schüler im Unterricht selbst zu.

3. Die Nutzung von Korpora im Klassenraum

Seit geraumer Zeit werden von Korpuslinguisten Vorschläge und Szenarien für den Einsatz von Korpora im fremdsprachlichen Unterricht und für die Nutzung von Korpora durch die Lernenden selbst entwickelt. Ein wesentlicher Grundgedanke ist dabei, dass Schülerinnen und Schüler durch die selbstständige Analyse von Korpusdaten die Zielsprache autonom und induktiv entdecken sollen. Ausgehend von entsprechenden Konzepten wie Widdowsons (1990) *learning as discovery* entwickelten Johns & King (1991) bereits in den frühen 1990er Jahren verschiedene konkordanzbasierte Übungen und Techniken des korpusbasierten entdeckenden Lernens, die sie mit Hilfe des Begriffs *Data-driven Learning* (DDL) zusammenfassten:

[Data-driven learning is] the use in the classroom of computer-generated concordances to get students to explore regularities of patterning in the target language, and the development of activities and exercises based on concordance output (Johns & King 1991: iii).

Es gibt inzwischen eine Vielzahl von Studien, die über DDL-Arbeitsweisen und DDL-Übungen berichten. Grundsätzlich lassen sich DDL-Aktivitäten auf einem Gradienten zunehmender Lernerzentriertheit und Lernerautonomie anordnen. Im Folgenden soll entlang dieses Gradienten anhand einiger ausgewählter Beispiele die Bandbreite möglicher konkordanzgestützter Aktivitäten im Englischunterricht verdeutlicht werden.

In Abbildung 1 wird ein Beispiel für eine geschlossene konkordanzbasierte Übungsaufgabe gegeben. Es handelt sich um eine Konkordanz, die von der Lehrperson selbst erstellt wurde, wobei das *node word* in der Mitte der Konkordanz durch eine Leerstelle ersetzt wurde (vgl. Flowerdew 2001: 368). Das Material soll Schülerinnen und Schüler dazu animieren, in authentischen Beispielsätzen Synonyme zum Adjektiv *important*, das von deutschen Lernern des Engli-

5 Diesen Rest nennt Mindt (2002: 211) im Übrigen "the compost of language", der untypischen und idiosynkratischen Sprachgebrauch ebenso wie beginnenden Sprachwandel umfasst.

schen übermäßig oft verwendet wird, in idiomatischen Adjektiv-Nomen-Kollokationen zu verwenden (z.B. *significant consequence, crucial factor*).

Alternative word list: critical / crucial / major / serious / significant / vital			
1.	nd imagine. This is the first and most	advantage of science and tech
2.	nimal and human lives. One of the most	consequences of pollution is the
3.	shown that they could work hard, take	decisions, and become involved
4.	ties seem to ignore completely another	factor people are slaves to:
5.	efficiency seem to be some of the all	factors in our hypersophistica
6.	television set. This has of course a(n)	impact on our mind: we are, ev
7.	e cause of women." This raises an	issue: according to me people
8.	g the environment. Air-pollution is a(n)	kind of pollution. Without be
9.	e in a world where television plays a(n)	part and it is highly probable
10.	ir reintegration in society. This is a(n)	point because in sane and c
11.	itely answer this question. A third	problem is the pollution due to
12.	the choice of the programmes, the most	question seems to be: Does te
13.	s in so critical a situation. The most	reason is pollution. Gas from
14.	concerned, it is the object of a very	requirement in the world of ad
15.	he influence of the media also plays a(n)	role. According to the amoun

Abbildung 1: DDL-Aktivität zur Einübung von Synonymen zu *important* in idiomatischen Adjektiv-Nomen-Kollokationen (vgl. Flowerdew 2001: 368)

Man kann sich auf der Grundlage des Beispielmaterials in Abbildung 1 sehr viele weitere Bereiche vorstellen, in denen Schülerinnen und Schüler auf eine ähnliche Weise einen korrekten und/oder idiomatische(re)n Gebrauch der Zielsprache einüben können, z.B. im Hinblick auf viele andere Kollokationen und lexikogrammatistische *patterns*, den korrekten Umgang mit dem englischen Tempus- und Aspektsystem, den Gebrauch von Artikeln mit zählbaren und nicht-zählbaren Substantiven (*count nouns vs. mass nouns*).

Während das in Abbildung 1 dargestellte DDL-Übungsmaterial von der Lehrperson selbst erstellt wurde, ohne dass Schülerinnen und Schüler selbst einen direkten Zugang zu Korpusdaten erhielten, veranschaulicht Abbildung 2 eine im Vergleich hierzu deutlich lernerzentriertere DDL-Aktivität (vgl. Gavioli 2001: 120). Es geht abermals um die idiomatische Verwendung von bedeutungsähnlichen Adjektiven, in diesem Fall *attractive, good, handsome, pretty, successful* und *tall*. Allerdings sollen in dieser DDL-Übung die Schülerinnen und Schüler selbstständig ein kleines Korpus aus Kontaktanzeigen mit Hilfe von

korpuslinguistischer *software* im Hinblick auf die verschiedenen Kollokationen und *patterns* der genannten Adjektive untersuchen.⁶

WORKSHEET

handsome good attractive pretty tall successful

- In pairs, choose three of these adjectives and have a look at their concordances.
 - Which words do they most frequently occur with?
 - Do they occur as part of series of adjectives?
- Discuss your findings with other pairs, then with the rest of the class.

Abbildung 2: DDL-Aktivität zur Erfassung der typischen Gebrauchskontexte von bedeutungsähnlichen Adjektiven (vgl. Gavioli 2001: 120)

Die Durchführung eines solchen korpusbasierten "Mini-Forschungsprojekts" durch Schülerinnen und Schüler setzt voraus, dass die Fremdsprachenlerner zumindest die Grundfunktionen korpuslinguistischer Software (so etwa die Erstellung von Konkordanzen, das Sortieren von Konkordanzzeilen, die Verwendung von so genannten *wildcards* zur Zulassung von optionalen Wörtern zwischen zwei Suchwörtern) kennen und kompetent nutzen können. Es zeigt sich, dass die umfassende *corpus literacy* des Lehrers (vgl. Abschnitt 2.2) auch deshalb eine Grundvoraussetzung für die stärkere Nutzung von Korpora im Fremdsprachenunterricht ist, weil es die Lehrperson ist, die Schülerinnen und Schülern korpuslinguistische Kompetenzen vermitteln muss. Haben sich Schülerinnen und Schüler diese Kompetenzen einmal angeeignet, können korpusbasierte "Mini-Forschungsprojekte" wie in Abbildung 2 veranschaulicht zu einer Vielzahl von Fragestellungen zum englischen Sprachgebrauch durchgeführt werden. So zeigt Kreyer (2007) am Beispiel der Verwendung der beiden Adjektive *older* und *elder*, wie mit Hilfe des *online* frei zugänglichen Konkordanzprogramms *Web Concordancer* (unter <http://vlc.polyu.edu.hk/concordance/WWWConcappE.htm>) Schülerinnen und Schüler auf verschiedenen Leistungsstufen die Gebrauchsunterschiede selbst erfassen können und wie diese korpusbasierten Erkenntnisse anschließend – ähnlich wie in Abbildung 1 gezeigt – über korpusbasierte Materialien eingeübt werden können. Mit Hilfe von Korpora lassen sich aber auch sehr schnell Fragen zum englischen Sprachgebrauch beantworten, die

⁶ Für die Analyse der genannten Adjektive ist ein genrespezifisches Korpus aus Kontaktanzeigen gut geeignet. Für viele andere Übungen wäre es sicherlich sinnvoller, genreübergreifende allgemeinsprachliche Korpora zu verwenden.

sich der individuellen Sprecherintuition (insbesondere der Lehrperson) entziehen bzw. anderweitig (etwa mit Hilfe von Wörterbüchern und Grammatiken) nur schwer beantwortet werden können. Hierzu zählen zum Beispiel die folgenden Fragen, die allesamt im Rahmen von sprachwissenschaftlichen Lehrveranstaltungen von Studierenden der Anglistik selbst aufgebracht wurden, weil sie hierzu keine Angaben in entsprechenden Nachschlagewerken gefunden hatten, vgl. a)–c), oder weil sie in englischsprachigen Fernsehsendungen auf für sie ungewöhnliche Formen gestoßen waren, vgl. d)–e):

- a) Verwendet man *in light of* oder *in the light of*? Wenn beide Varianten der komplexen Präposition vorkommen, welche Form ist häufiger/idiomatischer?
- b) Welche Adjektive sind typische Kollokate der Nomen *concern*, *issue* und *problem*?
- c) Welche der folgenden bedeutungsähnlichen Ausdrücke sind idiomatisch: *absolutely impossible*, *clearly impossible*, *fully impossible*, *totally impossible*?
- d) Ist der Satz *he's so not interested in my car* möglich? Nur im amerikanischen Englisch? Nur im gesprochenen (amerikanischen) Englisch?
- e) Ist die doppelte Verwendung des Verbs in *the thing is is that...* ein Fehler oder eine grammatisch wohlgeformte Konstruktion?

Das selbstständige Arbeiten mit Korpora kann im Sinne einer noch stärkeren Lernerautonomie gestaltet werden, indem man die oben genannten "Mini-Forschungsprojekte" zu größeren Analysen des Fremdsprachengebrauchs, die von den Lernern selbst durchgeführt werden, ausbaut. Im fortgeschrittenen Englischunterricht in der gymnasialen Oberstufe bieten die in den meisten Bundesländern obligatorischen Facharbeiten einen ausgesprochen geeigneten Kontext für eine wissenschaftliche Auseinandersetzung mit (fremd-)sprachlichen Phänomenen. Es wäre grundsätzlich wünschenswert, wenn der Themenkreis für solche Facharbeiten, die derzeit vor allem literaturwissenschaftliche, landeskundliche und historische Themen beleuchten, systematisch um empirische Analysen der Fremdsprache selbst erweitert würde – linguistische Korpora bieten hierfür hervorragende Datengrundlagen.

Sprachbezogene Facharbeiten können als ein besonderes Beispiel für die vielleicht selbstständigste Form des lernerzentrierten Arbeitens mit Korpora angesehen werden, und zwar das *learner-centred corpus browsing*, das bereits an anderer Stelle wie folgt beschrieben wurde:

- immer dann, wenn Schülerinnen und Schüler – aus welchen Gründen auch immer – ein besonderes Interesse an (oder auch besondere Probleme mit) einer sprachlichen Form haben, nutzen sie Korpusdaten, um sich authentische Gebrauchsinformationen zu verschaffen;
- zu diesem Zweck haben Schülerinnen und Schüler ständig die Möglichkeit, bei Bedarf auf Korpusdaten zurückzugreifen;
- die aus den Daten gewonnenen Erkenntnisse werden bei Bedarf mit der Lehrperson erörtert und gegebenenfalls vor der gesamten Lerngruppe präsentiert;
- die Funktion der Lehrperson besteht vor allem darin, auf diese selbstständige, lernerseitige Konsultation von Korpora vorzubereiten und stets bei Bedarf als Ratgeber zur Verfügung zu stehen (so z.B. bei Textverständnis- und Interpretationsproblemen);
- die Schülerinnen und Schüler sollten immer wieder dazu ermuntert werden, die Ergebnisse ihrer korpusbasierten Aktivitäten zu dokumentieren (denkbar ist z.B. auch eine systematische, lerngruppenspezifische Mappe mit den wesentlichen Ergebnissen) und in ihre eigene Sprachproduktion einfließen zu lassen;
- die Schülerinnen und Schüler sollten ebenso angeregt werden, bei der Analyse einer sprachlichen Form im Korpus auf andere, möglicherweise interessante Phänomene im Kontext zu achten und diese gegebenenfalls auch näher zu betrachten (Mukherjee 2002: 168).

Dieses Konzept des *learner-centred corpus browsing* basiert auf Bernardinis (2000, 2004) Ansatz eines *serendipitous corpus browsing*, vermeidet aber die völlig unstrukturierte Konfrontation von Lernern mit Korpusdaten, unter anderem durch die durchgehende Beratung und Hilfestellung durch die Lehrperson sowie durch die möglicherweise im Rahmen von korpusbasierten Facharbeiten klare Vorgabe von Ausgangsthemen für die Korpusexploration. Diese freie und umfassende DDL-Aktivität setzt allerdings ein hohes Maß an korpuslinguistischer Expertise auf Seiten der Lehrperson, Vorerfahrungen im Umgang mit Korpora und korpuslinguistischen Methoden auf Seiten der Lerner sowie einen angemessenen Zugang zu einer reichhaltigen Korpusumgebung (im Sinne eines *corpus-enriched classroom*) voraus.

Die genannten Beispiele dienen dazu, den Gradienten an möglichen DDL-Aktivitäten im Klassenraum von weitgehend geschlossenen Übungsaufgaben bis zu sehr freien Projektarbeiten abzustecken. Viele weitere DDL-Aktivitäten sind in der angewandten Korpuslinguistik inzwischen entwickelt worden (vgl. z.B. Aston 2001, Ghadessy, Henry & Roseberry 2001, Mukherjee 2002, Aston, Bernardini & Stewart 2004, Sinclair 2004). Auch gibt es darüber hinaus zahlreiche gut geeignete und frei zugängliche DDL-Seiten im Internet, so zum Beispiel T. Cobbs *Compleat Lexical Tutor* (unter <<http://www.lextutor.de>>).

Korpusbasiertes Fremdsprachenlernen muss nicht notwendigerweise den bislang geschilderten DDL-Szenarien entsprechen und auf Konkordanzanalysen oder konkordanzgestützten Übungen beruhen, sondern kann auch von Wortlisten (von Korpora) oder auch von der Analyse ganzer Texte ausgehen. Als ein im Englischunterricht erfolgreich umgesetztes Beispiel für eine textbasierte Metho-

de sei hier vor allem das von Henry & Roseberry (2001) beschriebene und einflussreiche Konzept des korpusbasierten *genre teaching* genannt. Die Methode verbindet einen textlinguistisch motivierten ersten analytischen Schritt mit einem auf die Schulung der Schreibkompetenz gerichteten, produktiven zweiten Schritt und nutzt dabei korpuslinguistische Ressourcen und Methoden:

- Im ersten Schritt werden Ganztexte eines genrespezifischen Korpus (z.B. aus wissenschaftlichen Aufsätzen bestehend) von den Lernern im Hinblick auf ihre typische Textstruktur (im Sinne von *textual moves*) hin untersucht: was sind die typischen Textbausteine und ihre kommunikativen Funktionen? Für jeden *move* (z.B. die *conclusion*) werden anschließend typische Versprachlichungsmittel, d.h. typische Ausdrücke und Wendungen, mit denen die kommunikative Funktion des *move* gut erfüllt werden kann, identifiziert (z.B. *to conclude...*, *in conclusion...*).
- Im zweiten Schritt produzieren die Lerner neue Texte des gleichen Genres, allerdings mit einer veränderten Themenstellung, wobei sie die im ersten Schritt beschriebene Textstruktur und möglichst viele der für jeden *move* typischen Ausdrücke und Wendungen als Gerüst übernehmen.

Rohrbach (2003) zeigt, wie diese Methode im Grundsatz bereits im Englischunterricht der neunten Klasse eines Gymnasiums, d.h. im vierten Lernjahr, eingesetzt werden kann. Er nutzt dabei die über das Internet frei zugänglichen Texte des Yorkshire Tourist Board, um ein *web-derived* DIY-Korpus mit Touristenbroschüren zu Yorkshire zu erstellen. Gemäß der von Henry & Roseberry (2001) vorgeschlagenen Methode analysieren die Schülerinnen und Schüler zunächst die Textstruktur der Broschüren im Hinblick auf die charakteristischen *moves*, ihre kommunikativen Funktionen sowie die in ihnen typischerweise vorkommenden Ausdrücke und Wendungen. Im Folgenden werden zur Veranschaulichung einige der in diesem Projekt identifizierten *moves* und ihre Funktionen genannt; für den fünften *move* werden beispielhaft einige typische Versprachlichungen genannt (vgl. Rohrbach 2003):

- Move 1: Promotion of the general character of an area
- Move 2: Presentation of fascinating history
- Move 3: Praise of beautiful landscape and natural sights
- Move 4: Recommendation of cultural assets
 - a) Museums
 - b) Exhibitions and events
- Move 5: Recommendation of night life and amusements
 - a) Clubs, Pubs etc.
 - *If clubbing means anything to you, X means everything to you.*

- *Don't miss out on X - the chance to enjoy...*
- *The more adventurous might want to try ...*
- b) Shopping
 - *Not to mention serious shopping ...*
 - *...including the only XY shop outside London.*
 - *The town has a... and is host to...*
 - *...all found in a pedestrianised city centre*
 - *For shoppers XY offers many choices*
 - *Of course there is...*
 - *For those who love to shop, XY offers a first-class opportunity to browse in unrivalled shopping facilities, including...*
 - *a weekly market is held every...*
- Move 6: Famous places or persons from movies and books
[...]

Auf der Grundlage der Ergebnisse der ersten Phase schrieben die Schülerinnen und Schüler anschließend eigene Touristenbroschüren, in denen sie die gleiche Textstruktur sowie möglichst viele der *move*-spezifischen Ausdrücke und Wendungen aus dem muttersprachlichen Korpus übernehmen sollten. Allerdings war nun das Thema nicht mehr Yorkshire, sondern ihre eigene Heimatstadt: Göttingen. Der folgende Text, der von einer Schülerin erstellt wurde, veranschaulicht trotz mancher Fehler, wie mit Hilfe der Methode des korpusbasierten *genre teaching* die genrespezifische Schreibkompetenz von Englischlernern sehr gut geschult werden kann und wie Lernertexte entstehen, die einerseits kreativ-innovativ ein neues Thema behandeln, andererseits aber die in der Zielkultur konventionalisierte Textstruktur sowie die üblichen idiomatischen Versprachlichungen beinhalten.

Situated in the south of Niedersachsen, Göttingen is famous for its university and is often called: university town Göttingen. Famous poets like Theodor Heuss [sic!] or Heinrich Heine described this "small city" as a small global city with an international atmosphere. In the town centre is the old marketplace with the old town hall and the famous "Gänselesel" fountain.

Built over 1,000 years ago as the village "Gutingi", Göttingen is home to a lot of historic churches and the romantic ruin of the Plesse Castle. A small piece of the city wall and a defence tower can also be visited.

Don't forget to enjoy the beautiful countryside around Göttingen, with its wooded hills and footpaths. The floral beauty and also the small villages in the region with their farms are breathtaking.

Discover the "Städtische Museum" with a great collection of musicinstruments. In addition there is the "Archäologisches Institut", which shows the artcollection of the university, the museum for Geologie and Paläontologie, the Völkerkundliche Sammlung and the Zoologisches Museum, presenting an exhibition of nature and animals with changing special exhibitions.

Don't miss out on Göttingen's nightlife. The club- and discoscene is very famous in the region. Many people from other towns come here to enjoy the entertainment. A variety of

bars, restaurants, cafés, discos, and nightclubs offers something for everyone. From business class to primitive pubs [sic!] for students.

If you want to stay here you've got wide range of accommodation to choose from: a cheap youth hostel or a noble hotel. If you want to have a wellness/sport hotel, the "Freizeit Inn" is a must. It's the third best wellness hotel in Germany. To sum it up, Göttingen is worth a trip (Rohrbach 2003: 388).

Das Beispiel des korpusbasierten *genre teaching* zeigt im Übrigen, dass korpuslinguistisch fundierte Methoden nicht nur im fortgeschrittenen Englischunterricht eingesetzt werden können, sondern, *mutatis mutandis*, auch bereits sehr viel früher sinnvoll zum Einsatz kommen können, und dass Fremdsprachenlerner durch das strukturierte Arbeiten mit authentischen muttersprachlichen Texten insbesondere ihre Idiomatizität im Gebrauch der Fremdsprache gezielt steigern können. Auf diese Weise entstehen Lernertexte, die sehr viel stärker an die (muttersprachliche) Zielnorm heranreichen.

Seit den 1990er Jahren widmet sich die anglistische Korpuslinguistik auch der Kompilation und Analyse von Korpora, die Lernertexte enthalten. Auf diese Weise ist auch der *output* von Lernern des Englischen zum Beschreibungsgegenstand der Korpuslinguistik geworden. Die Methoden und Erkenntnisse der Teildisziplin der Lernerkorpuslinguistik, auf die ich im folgenden Abschnitt eingehen möchte, sind von unmittelbarer Relevanz für die Fremdsprachenforschung und Fremdsprachendidaktik.

4. Der Lerner im Blickpunkt der Korpuslinguistik: Lernerkorpuslinguistik und Fremdsprachenforschung

Lernerkorpora werden von Granger (2002) wie folgt definiert:

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 homogeneous way and documented as to their origin and provenance (Granger 2002: 7).

Neben weiteren Studien belegen unter anderem Meunier (2002), Granger (2004) und Nesselhauf (2004) und Mukherjee (2007), dass die Analyse von Lernerkorpora wichtige Ergebnisse für die Fremdsprachendidaktik bereithält, zum Beispiel weil Lernerkorpora eine empirisch abgesicherte Basis für die quantitative Beschreibung häufig auftretender Fehler in den unterschiedlichen Stadien des Lehr- und Lernprozesses darstellen.

Das größte Lernerkorpus des Englischen, das bereits kompiliert worden ist, ist das *International Corpus of Learner English* (ICLE, vgl. Granger, Dagneaux

& Meunier 2002, Nesselhauf 2006). Es umfasst etwa zwei Millionen Wörter aus argumentativen Texten, die von fortgeschrittenen Englischlernern mit elf unterschiedlichen Muttersprachen (z.B. deutsch, französisch, schwedisch) in vergleichbaren standardisierten Situationen geschrieben wurden. Das gesprochene Pendant zu ICLE ist die *Louvain International Database of Spoken English Interlanguage* (LINDSEI, vgl. De Cock, Granger & Petch-Tyson 2006) – ein Korpus, das wie ICLE verschiedene Komponenten je nach Muttersprache der Lerner enthält. Die in LINDSEI enthaltenen Daten sind ebenfalls in standardisierten Interviewsituationen erhoben worden. Die deutsche Komponente von ICLE enthält deutlich über 200.000 Wörter, die deutsche Komponente von LINDSEI knapp 100.000 Wörter. Sowohl ICLE als auch LINDSEI sind internationale Projekte, die von Sylviane Granger an der Université Catholique de Louvain, einem der weltweit führenden Zentren der Lernerkorpuslinguistik koordiniert werden. Auch vor und neben diesen Megaprojekten gab und gibt es weitere Lernerkorpusprojekte. So wurde an der Universität Gießen nicht nur die deutsche Komponente von LINDSEI erstellt (vgl. Brand & Kämmerer 2006), sondern im Rahmen eines pragmatischen Forschungsprojekts ein weiteres, ähnlich großes gesprochenes englisches Lernerkorpus, nämlich das bereits erwähnte *Giessen Long Beach Chaplin Corpus* (GLBCC, vgl. Jucker, Smith & Lüdge 2003), das mündliche Erzählungen sowie Konversationen zwischen jeweils zwei Studierenden über den Inhalt des Stummfilms *The Immigrant* (mit Charlie Chaplin) enthält.

ICLE und LINDSEI können als Referenzlernerkorpora bezeichnet werden, da sie vergleichsweise groß sind und so konzipiert wurden, dass sie einen *archetypal learner* (Granger 1998: 7) repräsentieren, d.h. einen prototypischen Fremdsprachenlerner mit einer durchschnittlichen Sprachkompetenz. Der Vergleich von Referenzlernerkorpora mit muttersprachlichen Vergleichsdaten kann neuartige quantitative und qualitative Einsichten in diejenigen Bereiche des Sprachgebrauchs geben, in denen Englischlerner sich bereits an die (muttersprachliche) Zielnorm angenähert haben bzw. noch mehr oder weniger weit von der Zielnorm abweichen.⁷ So untersucht Lorenz (1999) eine frühe Version der deutschen Komponente von ICLE im Hinblick auf *adjective intensification* (d.h. die Verstärkung von Adjektiven durch prämodifizierende Elemente wie Adverbien) und

7 Die Annahme bei solchen Vergleichen, die in der Lernerkorpuslinguistik oft gezogen werden, ist natürlich, dass der muttersprachliche Sprachgebrauch in einer anglophonen Bezugskultur (i.d.R. Großbritannien und/oder die USA) die für den Englischunterricht relevante Zielnorm darstellt. Während diese Position insbesondere innerhalb der Lernerkorpuslinguistik vertreten wird (vgl. z.B. Mukherjee 2005, Granger 2006), wird von denjenigen, die das Englische heute vorrangig als eine *lingua franca* in der Kommunikation zwischen Nichtmuttersprachlern (und Muttersprachlern) ansehen, eine Abkehr von der muttersprachlichen Zielnorm für den Englischunterricht empfohlen (vgl. z.B. Seidlhofer 2001, Jenkins 2004).

vergleicht seine Befunde mit einem muttersprachlichen Vergleichskorpus. Er kategorisiert die statistisch signifikanten quantitativen Abweichungen im lerner-sprachlichen Gebrauch von der muttersprachlichen Verwendung der *adjective intensification* in vier Gruppen (vgl. Lorenz 1999):

- a) *Overuse*: Englischlerner verwenden bestimmte Formen signifikant häufiger als Muttersprachler (z.B. die Kombination 'really + Adjektiv' wie in *really big, really important, really interesting*).
- b) *Underuse*: Englischlerner verwenden bestimmte Formen signifikant seltener als Muttersprachler (z.B. die Kombination 'particularly + Adjektiv' wie in *particularly difficult, particularly useful*).
- c) *Misuse*: Englischlerner verwenden existierende *intensifiers* falsch, z.B. aufgrund von Interferenzen aus dem Deutschen (z.B. die falsche Verwendung eines prämodifizierenden *maximizer* vor einem nicht steigerbaren Adjektiv wie in *?absolutely silly, ?*totally damaged*).
- d) *Learner-idiosyncratic forms*: Englischlerner verwenden nicht existierende Formen oder Strukturen in der Zielsprache (z.B. das inkorrekte Muster 'a + too + Adjektiv + Substantiv' wie in **a too complex problem*).

Lorenz' (1999) Studie repräsentiert die größer werdende Zahl korpuslinguistischer Studien, in denen Lernerkorpora und muttersprachliche Vergleichskorpora miteinander verglichen werden, um hieraus auch fremdsprachendidaktisch relevante Schlussfolgerungen, etwa im Hinblick auf die Optimierung von Lehrmaterialien, zu ziehen, wobei ein besonderer Schwerpunkt bislang auf der Lexikogrammatik liegt.⁸ So zeigt Nesselhauf (2005) in ihrer lernerkorpuslinguistischen Studie, in welchen Bereichen fortgeschrittene deutsche Lerner des Englischen noch deutlich von Muttersprachlern im Kollokationsgebrauch abweichen. Andere Arbeiten analysieren darüber hinaus auch die Sprache in Lehrwerkstexten und vergleichen diese mit muttersprachlichen Korpora, um zu erörtern, inwiefern die Sprache in Lehrwerken 'natürliches Englisch' widerspiegelt (vgl. z.B. Römers (2005) Studie zur englischen Verlaufsform).

Auch die bereits fertiggestellten LINDSEI-Komponenten wurden bereits für korpuslinguistische Untersuchungen des gesprochenen Englisch fortgeschrittener Englischlerner genutzt. Dabei spielen diejenigen Phänomene eine besondere Rolle, die für spontan gesprochene Sprache besonders charakteristisch sind und zu 'natürlicher Mündlichkeit' beitragen. So untersucht beispielsweise Götz (2007) typische Performanz- und Hesitationsphänomene wie Wiederholungen (etwa von Pronomen), gefüllte und ungefüllte Pausen und Verbkontraktionen.

⁸ In vielen Fällen können die Abweichungen in der Lernaltersprache auf einen unmittelbaren oder mittelbaren Einfluss der Erstsprache zurückgeführt werden. So können alle in den Gruppen a) bis d) genannten Beispiele durch lexikalische bzw. kollokationelle und syntaktische bzw. lexikogrammatistische Eigenschaften des Deutschen erklärt werden.

Sie stellt anhand eines quantitativen Vergleichs der deutschen Komponente von LINDSEI und einem muttersprachlichen Vergleichskorpus in mehreren Bereichen deutliche Unterschiede fest. Die qualitative Analyse der Befunde befördert fremdsprachendidaktisch sehr wichtige Grundeinsichten zu Tage. Beispielsweise genügt Muttersprachlern offenbar in spontan gesprochener Sprache die durch die Wiederholung eines Funktionswortes zu Beginn einer syntaktischen Einheit "erkaufte" Planungszeit, um den Rest des Syntagmas ohne weitere Hesitationen zu produzieren, während für Englischlerner diese Wiederholung nicht ausreicht und oftmals auch danach der Planungsdruck in der Fremdsprache so groß ist, dass weitere Hesitationen wie z.B. Pausen auftreten. Götz (2007) schlägt daher die gezielte Vermittlung von natürlich wirkenden Kompensationsstrategien zur Überbrückung der häufigeren Planungspausen bei Englischlernern vor, so z.B. *discourse markers* wie *like, I don't know, you know, so, sort of* und *kind of*.

Während Referenzlernerkorpora wie ICLE und LINDSEI sehr wertvolle Ressourcen für den abstrakten und supraindividuellen Vergleich von Lernaltersprache und Muttersprache darstellen, sind Lehrpersonen naturgemäß vor allem an dem *output* ihrer eigenen Schülerinnen und Schüler in ihrem eigenen Unterricht interessiert. Insofern ist es für Lehrpersonen aus didaktischen Gründen von besonderem Interesse, so genannte 'lokale' Lernerkorpora mit der Lernaltersprache ihrer eigenen Schülerinnen und Schüler zu erstellen: "FL [foreign language] pedagogy, and presumably any pedagogy, has to be local, designed for specific learners and settings" (Seidlhofer 2002: 220). Ein computerisiertes lokales Lernerkorpus bietet der Lehrperson eine neuartige Möglichkeit zur Analyse des Lernstandes ihrer Schülerinnen und Schüler, so etwa für eine systematische Fehleranalyse. Mukherjee & Rohrbach (2006) beschreiben ein lokales Lernerkorpus, das aus 32.000 Wörtern aus Klausurtexten besteht und am Hainberg-Gymnasium Göttingen kompiliert wurde. Dieses Korpus, das *Giessen-Göttingen Local Learner Corpus of English* (GGLLC), besteht aus zwei Teilen: (1) alle Arbeiten zu einer Klausur in einem Englisch-Grundkurs der Jahrgangsstufe 12; (2) alle Arbeiten zu einer Klausur im gleichen Englisch-Grundkurs im darauf folgenden Schuljahr, d.h. in der Klasse 13. Mukherjee & Rohrbach (2006) zeigen, wie dieses Korpus ohne größeren Aufwand erstellt werden konnte und wie es in drei unterschiedlichen Formaten abgespeichert wurde: (1) als ursprüngliche Texte, wie sie von den Lernern geschrieben wurden; (2) als Texte mit den Korrekturmarkierungen der Lehrperson; (3) als Texte mit den Korrekturmarkierungen und den Verbesserungen der Lehrperson. Somit kann ein lokales Lernerkorpus wie das GGLLC unter anderem für folgende Zwecke eingesetzt werden (vgl. Mukherjee & Rohrbach 2006):

- Mit Hilfe korpuslinguistischer Software wie zum Beispiel *WordSmith Tools* kann die Lehrperson den allgemeinen und themenspezifischen Wortschatz der Lernenden untersuchen.
- Das Korpus ermöglicht es grundsätzlich, den *output* eines einzelnen Lernalters zu analysieren und ihm eine individuelle Rückmeldung zu seinem Lernstand zu geben.
- Das Korpus erlaubt eine longitudinale Analyse der Lernprogression innerhalb der gesamten Lerngruppe und für jeden einzelnen Lerner.
- Das Korpus kann sowohl von der Lehrperson als auch von den Schülerinnen und Schülern untersucht werden. So kann beispielsweise jeder Lerner eine Konkordanz mit den eigenen Fehlern in einer Fehlerkategorie (z.B. alle Grammatikfehler) erstellen.

Mit der Erstellung von lokalen Lernerkorpora kann man die Lernerkorpuslinguistik sehr nah an die am eigentlichen Unterrichtsgeschehen Beteiligten, nämlich die Lehrerinnen und Lehrer sowie die Schülerinnen und Schüler, heranführen. Aus Sicht der Lehrperson werden korpuslinguistische Methoden damit sehr viel relevanter für den eigenen Unterrichtskontext. Aus Sicht der Lerner ist die Auseinandersetzung mit dem eigenen *output* und die Reflexion über den eigenen Fremdsprachengebrauch motivierender als die Analyse von Texten, die von anderen produziert wurden.

Die Lernerkorpusanalyse ist ohne Zweifel einer der Bereiche der Korpuslinguistik, der eine besondere Relevanz für das Lehren und Lernen von Fremdsprachen aufweist. Die Diskussion um Standards für verschiedene Kompetenzebenen im Fremdsprachenerwerb könnte sicherlich von einer (lerner-)korpusbasierten Beschreibung der jeweiligen Deskriptoren wie etwa im Gemeinsamen Europäischen Referenzrahmen (vgl. Council of Europe 2001) sehr profitieren. Für die Zukunft ist aus fremdsprachendidaktischer Sicht unter anderem die Erstellung und Analyse von lokalen Lernerkorpora – auch im Zusammenhang mit lehrerzentriertem *classroom action research* – ein sehr vielversprechender und noch auszubauender Forschungszweig. Aus angewandt-linguistischer Perspektive ergibt sich als ein Desiderat für die zukünftige Forschung der systematische und umfassende Vergleich der Lernaltersprachen von Lernern des Englischen mit unterschiedlichen Erstsprachen. Das Design von Referenzlernerkorpora wie ICLE und LINDSEI ermöglicht es zum Beispiel, die Lernaltersprachen von deutschen, französischen und schwedischen Lernern des Englischen zu vergleichen und Gemeinsamkeiten und Unterschiede im schriftlichen und mündlichen Sprachgebrauch, auch im Hinblick auf Fehler und Normabweichungen vom muttersprachlichen Gebrauch des Englischen, festzustellen und auf diese Weise eine Lernaltersprachentypologie zu erstellen. Dieser Forschungsbereich wird als *Contrastive Interlanguage Analysis* (CIA, vgl. Granger 2002) bezeichnet.

5. Schlussbemerkungen

Der vorliegende Aufsatz stellt einen Versuch dar, einen Überblick über die wesentlichen Entwicklungslinien der anglistischen Korpuslinguistik aus der Perspektive der Fremdsprachenforschung anzubieten und ebenso einige wichtige Perspektiven für die zukünftige Forschungsarbeit aufzuzeigen. Die Relevanz korpuslinguistischer Ressourcen und Methoden für die Fremdsprachendidaktik ist in vielen Bereichen unübersehbar, wie die vorhergehenden Abschnitte gezeigt haben, so etwa in der Erstellung von Lernerwörterbüchern, -grammatiken und Lehrwerken für den Fremdsprachenunterricht, in der Nutzung von Korpora durch Lehrpersonen und Schülerinnen und Schülern im Rahmen von Szenarien des *data-driven learning* (DDL) und bei der Analyse des tatsächlichen Lernstandes von Fremdsprachenlernern mit Hilfe von Lernerkorpora.

Aus methodischer Sicht bleibt allerdings festzuhalten, dass es trotz der großen Zahl an Vorschlägen für korpusbasierte Aktivitäten im Englischunterricht nach wie vor eine große Lücke zwischen dem, was die angewandte Korpuslinguistik anbieten kann, und dem, was die durchschnittliche Lehrperson im Unterricht mit Korpora tut (bzw. nicht tut), gibt. Diese Lücke kann mittel- und langfristig nur dann geschlossen werden, wenn Lehrpersonen systematisch eine grundlegende *corpus literacy* vermittelt bekommen und wenn verstärkt korpusbasierte Aktivitäten im Englischunterricht unter authentischen Klassenraumbedingungen und in Echtzeit durchgeführt und – wichtiger noch – sowohl aus Lehrer- als auch aus Lernerperspektive evaluiert werden. In einer Fallstudie zur Durchführung und Evaluation von DDL-Aktivitäten im universitären Kontext zeigen Götz & Mukherjee (2006), dass Fremdsprachenlerner nur dann einen Mehrwert in korpusbasierten Übungen und Projektphasen sehen, wenn sie das Gefühl haben, die notwendigen Methoden und Techniken der Korpusanalyse zu beherrschen. Wenn die Lerner eine solche Kompetenz bei sich selbst sehen, empfinden sie DDL-Aktivitäten in der Regel als motivierend und nützlich. Diese Einschätzung zeigt, dass der Fremdsprachenerwerbsprozess auch aus Lernerperspektive mit Hilfe von Korpora gewinnbringend und interessant gestaltet werden kann.

Eingang des revidierten Manuskripts 05.03.2008

Literaturverzeichnis

- Aston, Guy (Hrsg.) (2001), *Learning with Corpora*. Houston, TX: Athelstan.
 Aston, Guy; Bernardini, Silvia & Stewart, Dominic (Hrsg.) (2004), *Corpora and Language Learners*. Amsterdam: John Benjamins.

- Bernardini, Silvia (2000), Systematising serendipity: proposals for concordancing large corpora with language learners. In: Burnard, Lou & McEnery, Tony (Hrsg.) (2000), *Rethinking Language Pedagogy from a Corpus Perspective: Papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt a.M.: Peter Lang, 225-234.
- Bernardini, Silvia (2004), Corpora in the classroom: an overview and some reflections on future developments. In: Sinclair, John (Hrsg.) (2004), 15-36.
- Biber, Douglas (1993), Representativeness in corpus design. *Literary and Linguistic Computing* 8, 243-257.
- Biber, Douglas; Conrad, Susan & Reppen, Randi (1998), *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, Douglas; Johansson, Stig; Leech, Geoffrey; Conrad, Susan & Finegan, Edward (1999), *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Brand, Christiane & Kämmerer, Susanne (2006), The Louvain International Database of Spoken English Interlanguage (LINDSEI): compiling the German component. In: Braun, Sabine; Kohn, Kurt & Mukherjee, Joybrato (Hrsg.) (2006), 127-140.
- Braun, Sabine; Kohn, Kurt & Mukherjee, Joybrato (Hrsg.) (2006), *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Frankfurt a.M.: Peter Lang.
- Burnard, Lou (Hrsg.) (2000), *Reference Guide for the British National Corpus (World Edition)*, <<http://163.1.0.36/docs/userManual/urg.pdf>>.
- Burnard, Lou (Hrsg.) (2007), *Reference Guide for the British National Corpus (XML Edition)*, <<http://www.natcorp.ox.ac.uk/XMLedition/URG/>>.
- Clear, Jeremy (2000), Do you believe in grammar? In: Burnard, Lou & McEnery, Tony (Hrsg.) (2000), *Rethinking Language Pedagogy from a Corpus Perspective: Papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt a.M.: Peter Lang, 19-30.
- Collins COBUILD English Language Dictionary (1987). London: Collins.
- Collins COBUILD English Grammar (1990). London: Collins.
- Council of Europe (Hrsg.) (2001), *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- De Cock, Sylvie; Granger, Sylviane & Petch-Tyson, Stephanie (2006), *The Louvain International Database of Spoken English Interlanguage – LINDSEI*, <<http://cecl.fltr.ucl.ac.be/Cecl-Projects/Lindsei/lindsei.htm>>.
- Dretzke, Burkhard & Nester, Margaret I. (2001), *Student's Guide to False Friends, Old Friends and New Friends*. Berlin: Cornelsen.
- Fandrych, Christian & Tschirner, Erwin (2007), Korpuslinguistik und Deutsch als Fremdsprache: ein Perspektivenwechsel. *Deutsch als Fremdsprache* 44: 4, 193-256.
- Flowerdew, Lynne (2001), The exploitation of small learner corpora in EAP materials design. In: Ghadessy, Mohsen; Henry, Alex & Roseberry, Robert L. (Hrsg.) (2001), 363-379.
- Francis, W. Nelson (1982), Problems of assembling and computerizing large corpora. In: Johansson, Stig (Hrsg.) (1982), *Computer Corpora in English Language Research*. Bergen: Norwegian Computing Centre for the Humanities, 7-24.

- Francis, W. Nelson & Kučera, Henry (1964), *Manual of Information to Accompany 'A Standard Sample of Present-Day Edited American English, for Use with Digital Computers'*. Providence, RI: Department of Linguistics, Brown University.
- Gavioli, Laura (2001), The learner as researcher: introducing corpus concordancing in the classroom. In: Aston, Guy (Hrsg.) (2001), 108-137.
- Ghadessy, Mohsen; Henry, Alex & Roseberry, Robert L. (Hrsg.) (2001), *Small Corpus Studies and ELT: Theory and Practice*. Amsterdam: John Benjamins.
- Götz, Sandra (2007), Performanzphänomene in gesprochenem Lernerenglisch: eine korpusbasierte Pilotstudie. *Zeitschrift für Fremdsprachenforschung* 18: 1, 67-84.
- Götz, Sandra & Mukherjee, Joybrato (2006), Evaluation of data-driven learning in university teaching: a project report. In: Braun, Sabine; Kohn, Kurt & Mukherjee, Joybrato (2006), 49-67.
- Granger, Sylviane (1998), The computer learner corpus: a versatile new source of data for SLA research. In: Granger, Sylviane (Hrsg.) (1998), *Learner English on Computer*. London: Longman, 3-18.
- Granger, Sylviane (2002), A bird's eye view of learner corpus research. In: Granger, Sylviane; Hung, Joseph & Petch-Tyson, Stephanie (Hrsg.) (2002), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, 3-33.
- Granger, Sylviane; Dagneaux, Estelle & Meunier, Fanny (Hrsg.) (2002), *International Corpus of Learner English*. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Granger, Sylviane (2004), Computer learner corpus research: current state and future prospects. In: Connor, Ulla & Upton, Thomas A. (Hrsg.) (2004), *Applied Corpus Linguistics: A Multi-dimensional Perspective*. Amsterdam: Rodopi, 123-145.
- Granger, Sylviane (2006), A plurilingual and multicultural perspective on learner corpus research. Paper presented at *Multilingualism & Applied Comparative Linguistics (MACL)*, Brussels, 8-10 February 2006.
- Gries, Stefan Th. (2008), *Quantitative Corpus Linguistics with R: A Practical Introduction*. New York: Routledge.
- Henry, Alex & Roseberry, Robert L. (2001), Using a small corpus to obtain data for teaching a genre. In: Ghadessy, Mohsen; Henry, Alex & Roseberry, Robert L. (2001), 93-133.
- Hunston, Susan (2002), *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Jenkins, Jennifer (2004), ELF at the gate: the position of English as a lingua franca. *The European English Messenger* 13: 2, 63-69.
- Johns, Tim & King, Philip (Hrsg.) (1991), *Classroom Concordancing*. Birmingham: University of Birmingham.
- Jucker, Andreas H.; Smith, Sara W. & Lüdge, Tanja (2003), Interactive aspects of vagueness in conversation. *Journal of Pragmatics* 35, 1737-1769.
- Kennedy, Graeme (1998), *An Introduction to Corpus Linguistics*. London: Longman.
- Kreyer, Rolf (2007), Das Korpus im Klassenzimmer. *Praxis Fremdsprachenunterricht* 6/2007, 17-21.

- Kreyer, Rolf (im Druck), Corpora in the classroom and beyond. In: Zhang, Felicia & Barber, Beth (Hrsg.) (im Druck), *Handbook of Research on Computer-Enhanced Language Acquisition and Learning*. Hershey, PA: IGI Global.
- Kreyer, Rolf & Mukherjee, Joybrato (2007), The style of pop song lyrics: a corpus-linguistic pilot study. *Anglia* 125: 1, 31-58.
- Lemnitzer, Lothar & Zinsmeister, Heike (2006), *Korpuslinguistik: Eine Einführung*. Tübingen: Gunter Narr.
- Lorenz, Gunter (1999), *Adjective Intensification - Learners versus Native Speakers: A Corpus Study of Argumentative Writing*. Amsterdam: Rodopi.
- Macmillan English Dictionary: School Edition for Advanced Learners* (2002). Hannover: Schroedel.
- McEnery, Tony; Xiao, Richard & Tono, Yukio (2006), *Corpus-based Language Studies: An Advanced Resource Book*. London: Routledge.
- Meunier, Fanny (2002), The pedagogical value of native and learner corpora in EFL grammar teaching. In: Granger, Sylviane; Hung, Joseph & Petch-Tyson, Stephanie (Hrsg.) (2002), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, 119-141.
- Meyer, Charles F. (2002), *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.
- Mindt, Dieter (2000), *An Empirical Grammar of the English Verb System*. Berlin: Cornelsen.
- Mindt, Dieter (2002), What is a grammatical rule? In: Breivik, Leiv Egil & Hasselgren, Angela (Hrsg.) (2002), *From the COLT's Mouth ... and others': Language Corpora Studies – in Honour of Anna-Brita Stenström*. Amsterdam: Rodopi, 197-212.
- Müller, Simone (2005), *Discourse Markers in Native and Non-native English Discourse*. Amsterdam: John Benjamins.
- Mukherjee, Joybrato (2002), *Korpuslinguistik und Englischunterricht: Eine Einführung*. Frankfurt a.M.: Peter Lang.
- Mukherjee, Joybrato (2004a), The state of the art in corpus linguistics: three book-length perspectives. *English Language and Linguistics* 8: 1, 103-119.
- Mukherjee, Joybrato (2004b), Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In: Connor, Ulla & Upton, Thomas A. (Hrsg.) (2004), *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam: Rodopi, 239-250.
- Mukherjee, Joybrato (2005), The native speaker is alive and kicking: linguistic and language-pedagogical perspectives. *Anglistik* 16: 2, 7-23.
- Mukherjee, Joybrato (2007), Exploring and annotating a spoken English learner corpus: a work-in-progress report. In: Volk-Birke, Sabine & Lippert, Julia (Hrsg.) (2007), *Anglistentag 2006 Halle: Proceedings*. Trier: WVT, 365-375.
- Mukherjee, Joybrato (in Vorbereitung), *Anglistische Korpuslinguistik: Eine Einführung*. Berlin: Erich Schmidt.
- Mukherjee, Joybrato & Rohrbach, Jan (2006), Rethinking applied corpus linguistics from a language-pedagogical perspective: new departures in learner corpus research. In: Kette-

- mann, Bernhard & Marko, Georg (2006), *Planing and Gluing Corpora: Inside the Applied Corpus Linguist's Workshop*. Frankfurt am Main: Peter Lang, 205-232.
- Nelson, Gerald C.; Wallis, Sean & Aarts, Bas (2002), *Exploring Natural Language: Working with the British Component of the International Corpus of English*. Amsterdam: John Benjamins.
- Nesselhauf, Nadja (2004), Learner corpora and their potential for language teaching. In: Sinclair, John (Hrsg.) (2004), 125-152.
- Nesselhauf, Nadja (2005), *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.
- Nesselhauf, Nadja (2006), Researching L2 production with ICLE. In: Braun, Sabine; Kohn, Kurt & Mukherjee, Joybrato (Hrsg.) (2006), 141-156.
- O'Keefe, Anne; McCarthy, Michael & Carter, Ronald (2007), *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.
- Pusch, Claus D. & Raible, Wolfgang (Hrsg.) (2002), *Romanistische Korpuslinguistik / Romance Corpus Linguistics*. Tübingen: Gunter Narr.
- Rieger, Burghard (1979), "Repräsentativität" – von der Unangemessenheit eines Begriffs zur Kennzeichnung eines Problems linguistischer Korpusbildung. In: Bergenholtz, Henning & Schaefer, Burkhard (Hrsg.) (1979), *Empirische Textwissenschaft: Aufbau und Auswertung von Text-Corpora*. Kronberg: Scriptor, 52-70.
- Römer, Ute (2005), *Progressives, Patterns, Pedagogy: A Corpus-driven Approach to English Forms, Functions, Contexts and Didactics*. Amsterdam: John Benjamins.
- Rohrbach, Jan (2003), "Don't miss out on Göttingen's nightlife": Genreproduktion im Englischunterricht. *Praxis des Neusprachlichen Unterrichts* 50, 381-389.
- Rundell, Michael & Stock, Penny (1992), The corpus revolution. *English Today* 8: 2, 9-14.
- Sampson, Geoffrey & McCarthy, Diana (Hrsg.) (2004), *Corpus Linguistics: Readings in a Widening Discipline*. London: Continuum.
- Schmied, Josef (1999), Applying contrastive corpora in modern contrastive grammars: the Chemnitz Internet Grammar of English. In: Hasselgård, Hilde & Oksefjell, Signe (Hrsg.) (1999), *Out of Corpora: Studies in Honour of Stig Johansson*. Rodopi, 21-30.
- Schmied, Josef & Haase, Christoph (2003), Grammatik lernen im Internet: die Chemnitz Internet Grammar. In: Keitel, Evelyne; Böhnke, Klaus & Wenz, Karin (Hrsg.) (2003), *Neue Medien im Alltag*. Lengerich: Pabst Science Publishers, 109-126.
- Scott, Michael (2005), *WordSmith Tools: Version 4.0*. Oxford: Oxford University Press.
- Seidlhofer, Barbara (2001), Closing a conceptual gap: the case for a description of English as a lingua franca. *International Journal of Applied Linguistics* 11: 2, 133-158.
- Seidlhofer, Barbara (2002), Pedagogy and local learner corpora: working with learning-driven data. In: Granger, Sylviane; Hung, Joseph & Petch-Tyson, Stephanie (Hrsg.) (2002), 213-234.
- Sinclair, John (1991), *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, John (Hrsg.) (2004), *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins.
- Stubbs, Michael (1993), British traditions in text analysis: from Firth to Sinclair. In: Baker, Mona; Francis, Gill & Tognini-Bonelli, Elena (Hrsg.) (1993) *Text and Technology: In Honour of John Sinclair*. Amsterdam: John Benjamins, 1-33.

- Stubbs, Michael (1995), Collocations and semantic profiles: on the cause of the trouble with quantitative studies. *Functions of Language* 2: 1, 23-55.
- Teubert, Wolfgang & Krishnamurthy, Ramesh (Hrsg.) (2007), *Corpus Linguistics: Critical Concepts in Linguistics*. 6 Bde. London: Routledge.
- Tribble, Christopher (2000), Practical uses for language corpora in ELT. In: Brett, Paul & Motteram, Gary (Hrsg.) (2000), *A Special Interest in Computers: Learning and Teaching with Information and Communications Technologies*. Whitstable, Kent: IATEFL, 31-41.
- Tsui, Amy B.M. (1996), The participant structures of TeleNex: a computer network for ESL teachers. *International Journal of Educational Telecommunications* 2(2/3), 171-197.
- Tsui, Amy B.M. (2004), What teachers have always wanted to know – and how corpora can help. In: Sinclair, John (Hrsg.) (2004), 39-61.
- Ungerer, Friedrich (1999), *Englische Grammatik heute*. Stuttgart: Ernst Klett.
- Walter, Maik & Grommes, Patrick (Hrsg.) (2008), *Fortgeschrittene Lernervarietäten: Korpuslinguistik und Zweitspracherwerbsforschung*. Tübingen: Niemeyer.
- Widdowson, Henry G. (1990), *Aspects of Language Teaching*. Oxford: Oxford University Press.

Französisch gegen Spanisch? Überlegungen aus Sicht der romanischen Mehrsprachigkeitsdidaktik

Daniela Caspari und Andrea Rössler¹

To learn French or to learn Spanish? That's the question for many students and their parents in Germany. To learn French and Spanish. That's the answer given in the following article. But why? The growing presence and prestige of Spanish in society and, consequently, as a foreign language in German schools is an obvious fact and French seems to be losing ground. However, the situation of French, which worries so many teachers, is not as bad as it may at first seem. With reference to the recommendation of the European Commission of Education to support individual and social multilingualism, our main purpose must be to enable every student to acquire at least three foreign languages at school. The advantages of combining French and Spanish (with English) are due to the fact that they belong to the same language family, thus facilitating language awareness, language learning awareness and the encouragement of learner autonomy. For those reasons, the article presents and compares three different types of intercomprehensive Romance language learning and describes the chances of their realization in the current German education system.

1. Einleitung

- "Man traut sich kaum zu sagen, dass man Französisch studiert. Die Spanisch-Studenten haben echt Oberwasser."
- "Das französische Kulturinstitut muss dafür sorgen, dass Spanisch nicht Französisch verdrängt, so wie in Frankreich Spanisch schon Deutsch als zweite Fremdsprache verdrängt hat."
- "An unserer Schule haben wir ab Jahrgangsstufe 7 jetzt sogar zwei Spanisch-Klassen. Das geht natürlich auf Kosten des Französischen."
- "Solange ich an dieser Schule Französisch-Fachleiter bin, werden wir kein Spanisch bekommen. Dafür Sorge ich."

Diese Stimmen – kürzlich in Berlin eingefangen – dürften so oder so ähnlich auch in anderen Bundesländern zu hören sein. Ihr emotional gefärbter Unterton zeigt, dass hier eine Entwicklung von der einen (französischen) Seite als bedrohlich empfunden wird und von der anderen (spanischen) Seite als motivierend: Das Spanische hat an den Schulen und Universitäten hohe Zuwachsraten zu

¹ Korrespondenzadresse: Prof. Dr. Daniela Caspari und Dr. Andrea Rössler, Freie Universität Berlin, Institut für Romanische Philologie, Didaktik der romanischen Sprachen und Literaturen, Habelschwerdter Allee 45, 14195 Berlin, Emails: caspari@zedat.fu-berlin.de, aroessle@zedat.fu-berlin.de