

Korpuslinguistik – Corpus linguistics – Les linguistiques de corpus: Unterschiedliche Stile der Einführung in diese Disziplin? Lektüererfahrungen eines der Einführung in die Korpuslinguistik bedürftigen weil blutigen Anfängers

Gottfried Kolde (Genf)

Abstract

In 2004, at the time of the seminar on German Corpus Linguistics held at Neuchatel, no introduction to this subdiscipline of German Linguistics was available. As a result, the author decided to describe, in his orally presented contribution to that seminar, his experiences in reading a French and an English book entitled *Les linguistiques de corpus* resp. *Corpus Linguistics*. In spite of the similarity of the titles, the differences in content and style were so great that even the dangerous notion of (different) national intellectual styles even came to mind. The French book turned out to be, in fact, not an introduction but a critical description of the actual state of art, whereas the English publication intends really to motivate and to inform the beginner. In the meantime just in time for the publication of the texts of 2004, there was published in 2006, the (nearly) ideal introduction to German Corpus Linguistics. It is reviewed in a short supplement to the text of 2004, as evidence for the rapid development of Corpus Linguistics as method and part of the linguistic description of natural languages and their use.

Der Vergleich der englischsprachigen Einführung in die *Corpus Linguistics* von Douglas Biber et al. (1998) mit einer französischsprachigen Darstellung betitelt *Les linguistiques de corpus* von Benoît Habert et al. (1997) war im Mai 2004 Gegenstand eines mündlichen Referats im Rahmen eines *Troisième Cycle* zur Korpuslinguistik des Deutschen in Neuchâtel, da eine deutschsprachige Einführung in diese Disziplin damals noch nicht auf dem Markt war. Inzwischen steht diejenige von Lothar Lemnitzer und Heike Zinsmeister (2006) zur Verfügung und es erschien sinnvoll, sie bei der Bearbeitung der damaligen Textvorlage zum Zweck ihrer Veröffentlichung in *linguistik online* in der Form eines Nachtrags zum im übrigen nur leicht veränderten Text von 2004 zu berücksichtigen.

Damit betreffen die hier dokumentierten Lektüererfahrungen folgende Veröffentlichungen: (Die drei eingehender behandelten Darstellungen sind durch Fettdruck hervorgehoben.)

Aijmer/Altenburg (1991)

Albert/Koster (2002)

Biber et al. (1998)

Brosius/Koschel (2001)

Habert et al. (1997)

Habert et al. (1998)

Kennedy (1998)

Lemnitzer/Zinsmeister (2006)

Lenders/Willée (1998)

Lenz (2000)

McEney/Wilson (2001)

Meyer (2002)

Pusch/Raible (2002)

Schlobinski (1996)

Der einzige Beitrag, den der Autor des folgenden Textes zu einem Troisième Cycle über die Korpuslinguistik des Deutschen leisten konnte, bestand darin, dass er über die Erfahrungen sprach, die er auf der Suche nach einer für ihn geeigneten Einführung in die Disziplin "Korpuslinguistik" gemacht hat. Und wenn man trotz intensiven Bemühungen keine Einführung findet, die in der Sprache verfasst ist und die sich ihre Beispiele und Probleme aus der Sprache holt, die einen vorab interessiert – im vorliegenden Falle dem Deutschen –, dann liegt es nahe, die Suche auf anderssprachige Einführung auszuweiten, ausgehend von der Idee, dass der Begriff der Korpuslinguistik übereinzelsprachlich definiert sei. Ob seine problemlose Übersetzbarkeit in andere Sprachen diese Idee hinreichend abstützt, bleibe dahingestellt. Denn nur allmählich wird es einem bewusst, welche zusätzlichen Perspektiven und welche Probleme man sich mit dieser Ausweitung des Suchbereichs auf alle möglichen Objektsprachen und nationalen Wissenschaftskulturen – wenn es denn solche gibt – einhandelt. Auch von diesen Perspektiven und Problemen wird im folgenden zu reden sein.

Dass der Autor sich im Untertitel als einen "der Einführung in die Korpuslinguistik bedürftigen, weil blutigen Anfänger in dieser Disziplin" charakterisiert, entspricht erstens den Tatsachen. Zweitens sind die Ausdrücke, mit denen er sich kennzeichnet, geeignet sich bewusst zu machen, was ein lexikalisch-semantisches und ein morphologisch-syntaktisches Analyseprogramm von der deutschen Sprache alles "wissen" müssten, wenn man sie obigen Untertitel als Teil eines elektronisch gespeicherten Korpus ohne menschliche Kontrolle und Nachbesserung analysieren lassen wollte: Dass ein *blutiger Anfänger* kein blutüberströmter, sondern ein wirklicher Anfänger ist, *blutig* hier also ebenso wie im Kontext von *Ernst*, *Laie* oder *Tränen* nicht als "Deskriptor", sondern als "Intensivierer" oder "Zentralitätsanzeiger" funktioniert, müsste ein semantisches Analyseprogramm erst einmal einer Kollokationenliste entnehmen. Und damit ein morphosyntaktisches Programm die Kasuszuweisung von *bedürftigen* und die syntaktische Struktur des ganzen Untertitels korrekt analysieren könnte, müsste es schon ziemlich detailliert sein. Schließlich könnte der Computer vielleicht zunächst versuchen, *weil* als Markierung einer Satzgrenze zu interpretieren, die folgende Konstituente folglich als abhängigen Satz zu annotieren und würde bei der Suche nach dem Verb in Endposition in die Brüche geraten. So etwa stellt sich jedenfalls der blutige Anfänger einige Probleme der Korpuslinguistik vor, nachdem er sich früher mit anderen Teildisziplinen der germanistischen Linguistik beschäftigt hat.

Da alle Teilnehmer an diesem *Troisième Cycle* zur Vorbereitung die 11 Seiten "Einführung in die Disziplin" lesen sollten, die Susanne Lenz ihrer *Studienbibliographie Korpuslinguistik* aus dem Jahre 2000 vorangestellt hat, erscheint es sinnvoll, sich zunächst des Gegenstandes der Korpuslinguistik in Anlehnung an diese "Einführung" zu vergewissern und sich sodann auf die Suche nach einer geeigneten selbständig erschienenen umfangreicheren Einführung in die so definierte Disziplin in dieser Bibliographie selbst zu begeben.

Im Anschluss an die Ausführungen auf S. 10 der Bibliographie von Susanne Lenz und die einschlägigen Abschnitte des "Wissenschaftlichen Arguments" unseres Seminars sei im folgenden unter "**Korpuslinguistik**" (= **KL**) die Zusammenstellung relativ grosser Textkorpora, ihre automatische Aufbereitung (von der Digitalisierung über die Segmentierung bis zur syntaktischen und semantischen Annotierung) sowie deren Nutzung zur Beantwortung linguistischer Fragen verstanden. Diese "Nutzung" kann – idealtypisch zugespitzt – zweierlei anstreben: Erstens die empirische Überprüfung von Hypothesen, die ihrerseits auf kompetenzlinguistischen Intuitionen des Forschers beruhen können, zweitens die weitere Entwicklung der linguistischen Theorie durch Hypothesenbildung auf der Grundlage einer systematischen Korpusanalyse. Im ersten Fall spricht man bekanntlich vom *corpus based approach*, vom Auffinden passender Belege oder dem Konsultationsparadigma, im zweiten vom *corpus driven approach* oder vom Analyseparadigma. Und ein **(Text)korpus** ist eine relativ große Menge vorgegebener (also nicht *ad hoc* hergestellter), aus praktischen Gründen meist schriftlicher,¹ digital gespeicherter und für verschiedene automatische Analysen präparierter Texte von gleicher oder verschiedener Textsorte bzw. Varietät aus meist einer Sprache (Ausnahme: Parallelkorpora für übersetzungswissenschaftliche Fragestellungen).

Aus obiger Explikation des Begriffs der KL kann man für eine Einführung in die KL drei anzusprechende Leserschaften und entsprechend anzustrebende Ziele ableiten: Es kann erstens um blutige Anfänger gehen. Die sind zunächst einmal zu motivieren, also davon zu überzeugen, dass sich der mit der KL verbundene Aufwand für sie lohnt. Dieser Aufwand kann heute noch für Philologen älterer Ausrichtung erheblich sein, denn Vertrautheit mit dem Computer und modernen digitalen Kommunikationsmitteln, die die jüngeren Kollegen heute schon aus ihrer Kindheit mitbringen, erleichtert zweifellos den Zugang, während sich Kompetenz- bzw. Intuitionslinguistik stets mit Papier und Bleistift betreiben lässt. Und Leser wie der Autor lassen sich vermutlich am ehesten dadurch für die KL motivieren, dass man ihnen die besonderen Vorzüge korpuslinguistischer Arbeiten an ihnen bekannten, aber auf der Basis sprachlicher Intuitionen nicht befriedigend beantwortbaren Fragen vorführt.

Auf einer zweiten Stufe der Initiation erwarten die Leser von einer Einführung in die KL, dass sie sie dazu ermutigt und befähigt, zur Beantwortung ihrer eigenen vermutlich zunächst mehr oder weniger kompetenzlinguistisch motivierten Fragen vorhandene Korpora mit den durch deren Annotierungen ermöglichten Analyseverfahren zu bearbeiten. Für diese zweite Leserschaft bedeutet es einen erheblichen Gewinn, dass man auch in der KL eine Tendenz beob-

¹ Dass sich die Beiträge zu dem von Pusch et al. (2002) herausgegebenen Sammelband ausdrücklich mit der (romanistischen) KL der gesprochenen Sprache beschäftigen, stellt also eine höchst lobenswerte Ausnahme dar.

achten kann, die in der kommerziellen Alltagsnutzung von Neuerungen der Telekommunikationsbranche noch deutlicher wird: Für reine Konsumenten wird die Nutzung digitaler Technik auch im Falle der KL durch die Entwicklung von Analyseprogrammen einfacher. Ein Beispiel: Die Entwicklung eines *www-online* Zugangs zum Mannheimer Korpus *CosmasII*. Außerdem wird immer mehr *Analyse-software* kommerziell angeboten.

Für solche Erleichterungen auf Stufe zwei dürfte dann aber die Stufe drei bezahlen: Die Entwicklung von Analyseverfahren, die für sehr spezielle linguistische Fragestellungen maßgeschneidert sind, dürfte so anspruchsvoll bleiben, dass es nur realistisch erscheint, hierfür eine weitere Disziplin anzusetzen, für die sich die Bezeichnung "Computerlinguistik" anbietet. Ob man den älteren Terminus "Linguistische Datenverarbeitung" synonym verwenden will oder für die Probleme reserviert, die noch näher bei der allgemeinen Informatik angesiedelt sind, ist im Kontext dieses korpuslinguistischen Seminars irrelevant. Das in Frankreich gebräuchliche Akronym *TALN* für "traitement automatique du langage naturel" bezeichnet alles, was die Informatiker machen und nicht die Linguisten. Und Informatiker sind Leute, die an einer höheren technisch-physikalisch-mathematischen Lehranstalt ausgebildet worden sind. Nicht nur bei der Konzeption der Korpora richten sich die Informatiker natürlich nach den Benutzerinteressen der Korpuslinguisten, sondern vor allem bei der Entscheidung darüber, welche Arten von Fragen mittels welcher Analyseprogramme sinnvoll an die Korpora gestellt werden können. Die Grenze zwischen den oben unterschiedenen Stufen 2 und 3 muss man sich im übrigen ziemlich fließend vorstellen: Durch die geschickte Kombination bereitgestellter Programmteile dürften künftig immer speziellere Phänomene erfass- und bearbeitbar werden.

Nun zur Suche in der Lenzschen Studienbibliographie. Einschlägig erscheint zunächst die Kennziffer *1.0 Einführende Literatur/Überblick*. Unter den 11 Einträgen finden sich unter (10.007) und (10.0010) zwei Bücher in deutscher Sprache, die ihrem Titel nach für unsere Zwecke geeignet sein könnten, während unter *6.2 Korpuslinguistik. Überblick* fast nur unselbständig erschienene Spezialliteratur aufgelistet wird. Aber wenn man sich diese zwei Veröffentlichungen, ihre Titel und Untertitel, die Inhaltsverzeichnisse und Einleitungen genauer ansieht, haben doch wohl auch sie nicht nur einen je etwas anderen Gegenstand, sondern verfolgen auch andere Ziele als wir hier in Neuenburg:

Der Titel des (akademischen) "Lehrbuchs" von Winfried Lenders und Gerd Willée (1998), *Linguistische Datenverarbeitung*, legt das Schwergewicht eindeutig auf die Computerlinguistik. Allerdings werden vor allem Probleme korpusbezogener Anwendungen erörtert: die Segmentierung oder "Tokenisierung" von Korpora, ihr grammatisches "Tagging" und die Strukturierung der Ketten in Bäumen und Netzen, weiterhin die Zusammenstellung lexikalischer Indices (Konkordanzen, Listen) als Basis für Wörterbücher und die statistische Auswertung der Korpora. Insofern vermittelt dieses Lehrbuch auch dem noch weitgehend uninformierten, aber bereits motivierten Leser nützliches korpuslinguistisches Wissen. Nur das letzte Kapitel über die *Umsetzung von Algorithmen in Programme* ist ganz der Entwicklung von Analysewerkzeugen gewidmet. Aber auch abgesehen von diesem Kapitel ist diese Darstellung keine Einführung in die KL im hier gesuchten Sinne: In bekannter deutscher Lehr-

buchtradition schenken die Autoren der Motivierung des Lesers für die KL kaum Aufmerksamkeit.

Den Gegenstand des anderen von Susanne Lenz unter "Einführungen" genannten deutschsprachigen Werks hat sein Autor Peter Schlobinski (1996) mit dem Titel *Empirische Sprachwissenschaft* charakterisiert. Nun wird niemand bezweifeln, dass die Korpuslinguistik eine empirische Wissenschaft bzw. eine empirische Methode wissenschaftlichen Arbeitens ist. Aber ebenso evident gibt es neben der Korpuslinguistik im oben präzisierten Sinne andere empirische Zugänge zu sprachlichen Phänomenen, etwa die Feldtechniken des Experiments, der Befragung, des Experteninterviews, der (teilnehmenden) Beobachtung. Diese liefern Daten(mengen), die man ebenso wie die Daten eines elektronisch gespeicherten Korpus vorgegebener Texte statistischen und qualitativen Analysen unterziehen kann, bis hin zu Inhalts-, Diskurs-, Konversations- oder ethnographischen Analysen. Um empirische Sprachwissenschaft in diesem Sinne geht es bei Peter Schlobinski und ebenso in zwei weiteren, nach Abschluss der Bibliographie von Susanne Lenz erschienenen Titeln: den *Methoden der empirischen Kommunikationsforschung. Eine Einführung* von Hans-Bernd Brosius und Friederike Koschel (2001) und dem *methodologischen Arbeitsbuch*, das Ruth Albert und Cor J. Koster 2002 unter dem Hauptitel *Empirie in Linguistik und Sprachlehrforschung* vorgelegt haben.

Damit sei die Suche nach einer geeigneten deutschsprachigen Einführung in die KL abgebrochen. Dass sie ergebnislos geblieben ist, mag daran liegen, dass sie zu früh abgebrochen wurde.² Aber vielleicht findet sich ja das, was gesucht wird, in einer anderen, etwa in der englischen oder der französischen Sprache, womit zunächst die Sprache der primären Redaktion des linguistischen Textes gemeint ist. Dass in der Regel auch die Korpora und die sprachlichen Phänomene, die in einer Einführung in "die" KL beispielhaft besprochen werden, derselben Sprache zugehören, ist vielleicht nur Ausdruck der Bequemlichkeit des Autors, könnte aber auch daran liegen, dass die KL des Deutschen nicht ganz genau das Gleiche ist wie die KL des Französischen und sich die Idee einer "allgemeinen KL" vielleicht zunächst im Falle der "englischen KL des Englischen" einstellt und sich letztlich eher der "Übereinzelsprachlichkeit" des Computers, also des Werkzeugs verdankt. Man kann sich zudem fragen, ob der Computer und bestimmte Arten von *software* als Werkzeuge wirklich für alle Typen von Objektsprachen gleich brauchbar sind.

Nutzen wird also der Griff nach einer anderssprachigen Einführung vor allem versprechen, soweit sie übereinzelsprachliche Aspekte und Probleme der KL behandelt. Damit ist die Frage nach der Position der KL zwischen den Polen "Allgemeine Linguistik" und "Linguistik einer bestimmten Einzelsprache" gestellt.

Dass hier gerade diese zwei weiteren Sprachen herangezogen werden, hat sehr einfache Gründe: Für das Englische spricht, dass die moderne Informatik und damit auch die KL im

² Im Mai 2004 musste der Verfasser noch mit dieser Möglichkeit rechnen. Dass auch Lemnitzer/Zinsmeister (2006) mehrfach auf das Fehlen einschlägiger Vorgängerarbeiten zum Deutschen hinweisen, verringert natürlich die Wahrscheinlichkeit mangelhaften bibliographischen Recherchierens.

englischsprachigen Raum und am Englischen als Gegenstand innerhalb und außerhalb des englischsprachigen Raumes seit langem und noch heute besonders intensiv betrieben wird. Dazu kommt, dass sich das Englische als *lingua franca* des interlingualen Wissenschaftsverkehrs in allen möglichen Disziplinen definitiv durchgesetzt hat. Schon aus diesen Gründen kann man erwarten, unter relativ zahlreichen und differenzierten englischen Einführungen in die KL wählen zu können, was sich denn auch voll bestätigt hat.³

Das Französische⁴ bietet sich umgekehrt aus Gründen der regionalen Organisation und Teilnehmerschaft dieses *Troisième Cycle* an: Der in der Westschweiz wirkende Germanist lebt selbst im französischen Kulturraum und in engem Kontakt mit der französischen Wissenschaftswelt und ist es gewohnt, sich mit deren Besonderheiten auseinander zu setzen. Die überraschenden Schwierigkeiten, auf die die Suche nach einer den englischen entsprechenden französischsprachigen Einführung in die KL stößt, verlangen eine Erklärung. Insbesondere muss man sich fragen, ob die Gründe die gleichen sind wie im Falle der deutschsprachigen KL. Wie erinnerlich haben wir die Suche nach einer geeigneten Einführung in diese Disziplin bereits oben ergebnislos abgebrochen. Die damit angesprochenen allgemeinen Fragen seien nun zunächst etwas weiter konkretisiert, bevor wir zu den speziellen Erfahrungen mit jenen Darstellungen der KL kommen, die wir der Erweiterung unseres Suchfeldes auf das Englische und das Französische als Beschreibungs- und Objektsprache verdanken.

Mit der Erweiterung der Perspektive über die deutsche KL hinaus kommen zwei Aspekte in den Blick, die den Autor seit längerem interessieren und die man mit den Stichwörtern (*nationaler*) *Wissenschaftsstil* und *interlinguale Sprachbarrieren* plakativ umreißen könnte: Dass Wissenschaft und ihre didaktische Vermittlung in England und Nordamerika nach recht ähnlichen Mustern betrieben würden, (etwas) anders hingegen in Frankreich und noch anders in den deutschsprachigen Ländern, dass man also so etwas wie verschiedene "intellektuelle Stile" oder Wissenschaftskulturen unterscheiden könne, die "irgendwie" mit unterschiedlichen, historisch gewachsenen und vermittelten Kommunikations- und Denkkulturen zusammenhängen, die man ihrerseits an den Nationalsprachen als praktischen Identifikationsinstanzen für intensiv interagierende soziale Gruppen festmachen könne – das ist eine ebenso verbreitete wie umstrittene, höchst spekulative, simplifizierende und vorurteilsträchtige Meinung,⁵ die ihre Nähe zur Sapir-Whorf-Hypothese nicht verbergen kann. Die reelle Chance einer Veri- oder Falsifizierung hätte sie wohl am ehesten in ihrer schwächsten Variante, einer Variante, die davon ausgeht, dass Wissenschaft (allgemein oder eine bestimmte Disziplin oder eine bestimmte intellektuelle Teilaktivität) in den verschiedenen Gesamtgesellschaften unterschiedliche Wertschätzung genieße und zudem auf unterschiedliche Weise (universitär) vermittelt werde. So heisst es immer wieder, englische, insbesondere US-amerikanische

³ Als Musterbeispiel für die Dominanz des Englischen in der KL kann Aijmer et al. (1991) gelten: 15 der 24 Beiträger lehren in einem englischsprachigen Land, 8 in Skandinavien, einer in den Niederlanden, zwei in Deutschland. Alle 24 sind Anglisten und schreiben ihre Beiträge auf Englisch.

⁴ Wie *ex negativo* aus vorstehender Fußnote folgt, lebt keiner der Beiträger zu Aijmer et al. (1991) in einem romanischsprachigen Land, ist romanischsprachig oder Romanist von Beruf. Vgl. auch den Anhang 1.

⁵ Vgl. hierzu die Beiträge von Robert B. Kaplan (1966), Johan Galtung (1985) und Michael Clyne (1987), die Kirsten Adamzik in ihrer *Kontrastiven Textologie* (Adamzik 2001: 17-20) vorstellt.

Forscher gäben sich besondere Mühe, ihre Ergebnisse dem Laien leicht verständlich zu vermitteln, während deutsche und französische Forscher aus vielleicht unterschiedlichen Gründen eher Schwerverständlichkeit pflegten (Verständlichkeit als Bringschuld des Fachmannes oder als Holschuld des Laien: vgl. Michael Clyne 1987: 238).

Wer solche Beobachtungen macht, geht letztlich von relativ geschlossenen (kultur- bzw. sprach)nationalen Wissenschaftstraditionen und durch ihre Sprache definierten Wissenschaftlergemeinschaften aus. Was nicht ausschließt, dass eine der Sprachen in Folge der Größe und des Ansehens ihrer Sprechergemeinschaft zur *lingua franca* der weltweiten Zusammenarbeit und Vernetzung einer Wissenschaften wird: Dass die 26 Beiträger zu Aijmer et al. 1991 einen homogenen Einführungstext geschaffen haben, obwohl sie aus neun verschiedenen Ländern kommen, kann man damit erklären, dass sie im gleichen "englischen" Wissenschaftsverständnis sozialisiert sind und arbeiten. Dass unter diesen 26 Korpuslinguisten hingegen kein Franzose oder anderer Romane ist, ist natürlich ein Zufall, aber einer, der gut zu den Lektüreerfahrungen passt, von denen hier die Rede sein wird.

Ein Schwachpunkt des Konstrukts kulturellen Wissenschaftsvermittlungsstile und deren eindeutiger Anbindung an je eine bestimmte natürliche Einzelsprache liegt darin, dass die (Meta)sprache eines wissenschaftlichen Textes gar nicht die Sprache zu sein braucht, mit der zusammen bzw. in der sein Verfasser seinen intellektuellen Stil erworben und in der er seine Erkenntnisse gewonnen bzw. in der er diese Erkenntnisse der wissenschaftlichen Öffentlichkeit (bzw. den "Einzuführenden") bekannt zu machen gelernt hat. Für die Bestätigung oder Widerlegung des Konzepts der sprachnationalen Wissenschaftsstile wären wohl auch alle linguistischen Texte von Bedeutung, deren Autor sie nicht in seiner eigenen Sprache, sondern in der der jeweiligen Forschergemeinschaft geschrieben hat, welch letztere oft auch die Objektsprache ist. Wie weiter oben schon erwähnt, ist zwar die Übereinstimmung dieser Sprachen in der heutigen KL (vielleicht überraschenderweise) die Regel. Aber das heißt: es gibt Ausnahmen, etwa unter den Beiträgern zu Pusch et al. (2002). In diesen Fällen müsste sich zeigen lassen, welcher Stil sich durchsetzt (der der Forschergemeinschaft ("Romanisten" romanischer Herkunft) oder der des Autors nichtromanischer Herkunft (also eines "*romantisant*"), oder ob der resultierende Text einfach stilistisch weniger oder gar nicht markiert ist (vgl. Anhang 1).

Falls sich etwa in einer deutschen und einer französischen Darstellung der KL nur ganz wenige anderssprachige Literaturverweise finden würden und man hieraus auf einen schwachen oder gar fehlenden interlingualen Informationsaustausch schließen würde, bliebe es doch grundsätzlich unentscheidbar, ob dieses Manko auf Sprachbarrieren im kruden Sinne fehlender Fremdsprachenkenntnisse bzw. -interessen der Wissenschaftler zurückzuführen wäre, auf als irritierend empfundene Gegensätze im wissenschaftlichen Arbeits- und Kommunikationsstil der anderssprachigen Kollegen oder darauf, dass man in deren Forschungsproblemen und -ergebnissen die eigenen nicht wiederfindet, etwa weil die betreffenden Sprachen zu weit von einander typologisch entfernt sind. Für die im französisch- und deutschsprachigen Raum betriebene KL ist die anglo-amerikanische KL zur Zeit zudem vielleicht einfach zu nützlich weil fortgeschritten, um zusätzlich noch die im kontinentaleuropäischen Nachbarland betrie-

bene KL zur Kenntnis nehmen zu können.⁶ Teilnehmende Beobachtung während vielen an einer "sehr französischen" Universität verbrachten Jahren sowie Erfahrungen anlässlich mancher deutsch-französischen linguistischen Fachtagung sprechen für eine Kombination dieser verschiedenen Ursachen des realen Mangels an interlingualer wissenschaftlicher Interaktion, den wir weiter unten werden konstatieren müssen.

Auf solche Fragen gerät man, wenn man die gesuchte Einführung in die KL einer bestimmten Sprache nicht findet und darum auf vorhandene Einführungen in die KL anderer Sprachen ausweichen will, dabei auf große Unterschiede im Gegenstands- und Zielverständnis der verschiedensprachigen Autoren stößt und diese mit Unterschieden in den nationalen Wissenschaftsvermittlungsstilen in Verbindung zu bringen versucht ist.

Nach diesen umständlichen und reichlich abstrakten Vorbemerkungen kommen wir nun endlich zu der Frage, ob sich unter den verfügbaren nicht-deutschsprachigen Einführungen in die KL eine, womöglich mehrere finden, deren Lektüre jemandem wie dem Referenten zu empfehlen ist, auch wenn die beispielhaft behandelten Probleme und sprachlichen Phänomene gerade nicht die Sprache betreffen, die den Leser vorab interessiert.

Bei den englischsprachigen ergab sich in der Tat die Qual der Wahl unter mehreren geeigneten Darstellungen, die sich u.a. darin voneinander unterschieden, dass die einen ihren Gegenstand im Titel kurz und bündig als *Corpus Linguistics* bezeichnen, andere genauer formulieren und eine Einführung in die *English Corpus Linguistics* versprechen, was nach dem oben Gesagten einen nicht unwichtigen Unterschied anzeigen könnte. Leider sagt dieser Unterschied in der Titelformulierung aber offenbar manchmal gar nichts über einen unterschiedlichen inhaltlichen Schwerpunkt des Textes. Die *English Corpus Linguistics* von Charles Meyer (2002) oder die unter dem gleichen Titel schon 1991 von Karin Aijmer und Bengt Altenberg organisierte "Einführungsfestschrift" könnten gar nicht ausschließlicher auf das Englische ausgerichtet sein als die schlussendlich zur genaueren Vorstellung ausgewählten *Corpus Linguistics. Investigating Language Structure and Use* von Douglas Biber, Susan Conrad und Randi Reppen, 1998 in der Reihe *Cambridge Approaches to Linguistics* in Cambridge (UK) erschienen. Ihre Verfasser lehren an der Northern Arizona University bzw. der Iowa State University. Dem Literaturverzeichnis zufolge hat Biber viel über die mediale Varianz (*gespr.* vs. *geschr.*) im amerikanischen Englisch gearbeitet und ist an der Entwicklung der *Longman Grammar of Spoken and Written English* beteiligt. Conrad hat ihren PhD über die US-englischen Fachsprachen der Biologie und der Geschichtswissenschaft geschrieben, Rieben über die US-englische Schülersprache.

In französischer Sprache fand sich nur ein einschlägiger⁷ Titel: *Les linguistiques de corpus* von Benoît Habert, Adeline Nazarenko und André Salem von 1997. Wie noch zu zeigen ist,

⁶ Johan Galtung (1985: 184) vermutet, dass generell die "industrialisierte intellektuelle Produktionsweise" des Sammelns und Analysierens großer Datenmengen mittels EDV dem "sachsonischen" intellektuellen Stil der Thesenproduktion entspreche und diesem darum als ein "ungeheuer wirkungsvolles Herrschaftsinstrument" diene.

⁷ Das ebenfalls von Benoît Habert mit zwei anderen Kollegen 1998 veröffentlichte Lehrbuch *De l'écrit au numérique. Constituer, normaliser et exploiter les corpus électroniques* behandelt ausführlich die Bereitstellung

wollen sie mit dem überraschenden "echte" Plural von "KL" (im Gegensatz zum englischen *plurale tantum Corpus Linguistics*) im Titel die herrschende, ihrer Meinung nach schädliche Vielfalt der konkurrierenden korpuslinguistischen Ansätze kritisieren. Diese nicht in der Lenzschen Bibliographie, aber im Wissenschaftlichen Argument dieses Seminars erwähnte Darstellung hat ebenso drei Autoren wie Biber et al. 1997, alle drei waren beim Erscheinen des Buches 1997 *maître de conférences* in Saint Cloud bzw. Paris. Benoît Habert hat nach Ausweis des Literaturverzeichnisses über die automatische syntaktische Analyse gearbeitet, Adeline Nazarenko über semantische Etikettierungen und André Salem über statistische Analyse.

Zunächst das Ergebnis der vergleichenden Lektüre von Biber et al. (1998) und Habert et al. (1997): Bei Biber et al. (1998) handelt es sich, dem Reihentitel *Cambridge Approaches to Linguistics* entsprechend, um eine echte "Einführung" in die KL, also eine Darstellung, die versucht, den Laien bzw. Anfänger mit Beispielen aus der eigenen Forschung davon zu überzeugen, dass die KL "Spaß macht" und sich "lohnt". Dies ist auch der Grund für ihre etwas ausführlichere Besprechung hier.⁸

Habert et al. (1997) gehört hingegen einer ganz anderen Textsorte an: Es ist ein kritischer Forschungsbericht, ein "panorama de travaux récents", eine Dokumentation des *state of the art* mit dem Ziel, die Kollegen dazu aufzurufen, die bestehenden, wohl nur zum Teil typisch französischen institutionellen Konflikte zwischen Informatikern und Linguisten, zwischen Theoretikern und Praktikern beizulegen sowie die theoretischen und technischen Divergenzen und Animositäten zu überwinden, die die Zusammenarbeit und damit den wissenschaftlichen Fortschritt hemmen. So fordern die Autoren, die Annotationsformate zu vereinheitlichen, um die für bestimmte Fragestellungen annotierten Korpora auch für andere Forschungszwecke wiederverwendbar zu machen. Die von den Autoren zitierten Beispiele aus den Forschungen der Kollegen mögen gelegentlich sogar eher eine abschreckende Wirkung auf den blutigen Anfänger ausüben, denn es werden vor allem Schwierigkeiten, Mängel und Forschungsdesiderate benannt. Schließlich erscheint die englisch-amerikanische KL als große Herausforderung für die noch unzureichend entwickelte französische Forschung. Entsprechend betonen die Autoren ausdrücklich (S.12, Anm.11), dass sie sich bemühen, terminologische Anglizismen zu vermeiden. Und man sollte eines nicht vergessen: Habert und seine Koautoren bezeichnen ihr Buch nirgends als eine Einführung (*introduction*)!

und Bearbeitung von Korpora und die Herstellung von *software* für deren Analyse, aber nicht deren Nutzung zur Beschreibung bestimmter sprachlicher Phänomene (Das *exploiter* des Untertitels meint u.a. das *Parsing*), ist also der der Computerlinguistik im oben eingeführten Sinne zuzuordnen. Die von Pusch et al. (2002) herausgegebene *Romanistische Korpuslinguistik* ist eine Sammlung von Forschungspapieren.

⁸ Auch Meyer (2002) ist ohne Zweifel eine überaus gelungene Einführung, die aber schon mehr Gewicht auf die technischen Detailprobleme und -informationen legt. Eine reizvolle Alternative wäre auch Aijmer et al. (1991), eine Festschrift für Jan Svartvik, deren 26 Beiträger, in der Mehrzahl bedeutende Vertreter der KL, in 20 Kapiteln alle Themen abhandeln, die in einer Einführung in die KL abgehandelt zu werden pflegen: in drei Beiträgen geht es um "goals and methods", in zweien um "corpus design and development", in 13 um Anwendungen ("exploration of corpora") und in einem um "prospects for the future". Mit der Wahl dieser "Einführung" hätte auch der blutige Anfänger die Chance, sogleich wirkliche Autoritäten kennenzulernen.

Dieser globale Leseindruck basiert auf Zitaten wie den folgenden:

- (1) "... we chose to write a book that reflects the reasons of our own enthusiasm for corpus-based research: this is a book about language as well as corpus linguistics, showing how empirical investigations of corpora can shed new light on previously untractable research questions in linguistics." (Biber et al.1998: IX)
- (2) "... this book will show you the wide range of language issues that can be addressed using the corpus-based approach. It will develop your skills as an educated and critical reader of corpus-based studies. We hope that it will also inspire you in developing your own reasearch questions about language use and in conducting your own corpus-based investigations to answer those questions." (Biber et al. 1998: 18)
- (3) "As we have shown, corpus based studies greatly expand our understanding of language and language use. Corpora and computer tools are increasing in availability dramatically, and as a result, the number of corpus-based studies is exploding. Thus we hope that the introduction presented here will encourage you to pursue this approach to linguistics, whether you proceed as a consumer or as an active researcher carrying out your own projects." (Biber et al. 1998: 240)
- (4) "Cet ouvrage présente un panorama de travaux récents dans le domaine du traitement automatique des textes..." (Habert et al. 1997: Klappentext, hinterer Einband, außen)
- (5) "... Or, l'étude des origines de ces travaux le montre, ce sont les discontinuités qui l'importent, ainsi que la diversité, voire l'éclatement, des horizons théoriques et des réalisations pratiques." (Habert et al. 1997: 8)
- (6) "Nous parlons de linguistiques de corpus au pluriel pour souligner cette diversité d'approches." (Habert et al. 1997: 15)
- (7) "Institutionnellement, en France, le TALN [= traitement automatique du langage naturel] et la linguistique relèvent de deux secteurs disciplinaires aux fonctionnements éloignés: entre ces domaines, les passerelles et les collaborations sont encore fragiles. Les formations autour du traitement automatique du langage, par exemple, relèvent dans l'immédiat d'un secteur ou de l'autre, mais pas d'une convergence des deux." (Habert et al. 1997, 215)
- (8) "Les linguistiques de corpus se révéleront fructueuses comme domaine de recherche si [...] s'affirment les collaborations soutenues entre linguistes et informaticiens. [...] Il y a probablement une chance historique à saisir: celle d'une coopération enfin fructueuse." (Habert et al. 1997: 217f.)

Der Textsortenunterschied ist mit den Händen zu greifen: Einführung vs. sehr kritischer Forschungsbericht. Dass es zumindest sehr viel einfacher ist, eine englische Einführung in die KL zu finden als eine französische, könnte man nun aber bereits als einen Hinweis auf bestehende Unterschiede in den Wissensvermittlungskulturen im oben ausgeführten Sinne interpretieren, und der am Ende dieses Textes als Anhang 2 wiedergegebene Versuch, die Makrostruktur der beiden Darstellungen in einer zweiseitigen Synopse sichtbar zu machen, lässt manche Unterschiede erkennen, die mehr oder weniger deutlich an das denken lassen, was man zu den nationalen Stilen der Wissenschaftsvermittlung geschrieben hat. So ist der Teil IV, *Methodological boxes* von Biber et al. (1998) von besonderem Interesse. Diese methodologischen Ratschläge **folgen** der *Conclusion (Summing up and looking ahead)* des 9. Kapitels.

Sie gehören also nicht eigentlich zur "ersten Einführung". Ganz anders bei *Habert et al.* (1997). Hier sind die thematisch allenfalls vergleichbaren Kapitel III.7 bis III.9 zur Konstituierung und Annotierung der Korpora und zur quantitativen Analyse **vor** der *Conclusion* positioniert, weil auch sie weniger Ratschläge oder Rezepte, als vielmehr kritische Kommentare bieten. Im Blick auf die oben vorgestellten drei aufeinander aufbauenden Ziele einer Einführung in die KL betonen Biber et al. 1998 in der *box 3* zum Beispiel ausdrücklich, dass für ernsthafte und originelle Forschung auch die dritte Stufe der Initiation in die KL erreicht werden müsse. Denn käufliche *concordancing packages* reichten nicht aus, der Forscher müsse sich seine speziellen Analysewerkzeuge selbst schaffen. Und man könne das. Denn:

- (9) "Fortunately (and contrary to many people's fears), programming for linguistic research questions does not require a special aptitude in computer science or mathematics [...] and it is simply a matter of learning how to use it [= a language that a computer understands]."

Allerdings:

"It is easier to learn to program in a course. [...] Therefore, a course in programming for linguistic analysis is likely to be most useful for you." (Biber et al. 1998: 256)

Bei Habert et al. (1997) würde man solch eine didaktisch gemeinte Kombination von Ermutigung und Ratschlag vergeblich suchen, denn diese Korpuslinguisten machen sich über die notwendigen und nicht selbstverständlichen Spezialqualifikationen ihrer Leser gar keine Gedanken, bzw. gehen stillschweigend davon aus, dass dieselben die linguistische Datenverarbeitung perfekt beherrschen. In ihrem "typisch französischen"⁹ Bemühen um intellektuelle Originalität und aparte, oft rhetorisch-stilistisch markierte Formulierungen lassen sie den Leser in der Regel allein in dem Bemühen, sie zu verstehen. Das gilt sogar für manche Kapitelüberschriften. So spalten sie einerseits den semantischen Aspekt auf die Kapitel I,3 und II,4 auf, die sie etwas präventiv mit "*I,3 Les ressources lexicales pour l'étiquetage sémantique*" und "*II,4 Des mots aux sens: sémantique en corpus*" betiteln und fassen andererseits letzteres Kapitel II,4 mit den folgenden II,5 (über diachrone Korpora) und II,6 (über mehrsprachige Korpora) unter dem dunkel-raummetaphorischen Titel *Dimensions transversales* zusammen. Das klingt bedeutsam, aber die Mühen, dieser Metapher einen verbindlichen Sinn zu verleihen, überlassen sie dem Leser. Damit dürfte zumindest einer der Gründe für die spezifischen Verständnisschwierigkeiten mancher Leser mit der "typisch gallischen" Wissenschaftssprache angesprochen sein. Es ist hier nicht der Ort, um die gelegentlich geäußerte Vermutung im Detail zu entwickeln und zu überprüfen, dass die ebenso oft festgestellte Schwerverständlichkeit des "teutonischen" Stils der Wissenschaftsvermittlung eher in dem "pedantischen" Bemühen begründet sei, die Komplexität des jeweiligen Gegenstandes in der sprachlichen Form abzubilden, und damit andere Gründe habe als im Fall des Französischen.

Kurz nur zu den interlingualen Kontakten, so wie sie in den Literaturverzeichnissen zum Ausdruck kommen. Diese Kontakte erweisen sich als hochgradig einseitig: Biber et al. (1998) zitieren nur englischsprachige Literatur, Habert et al. (1997) zu zwei Dritteln englischsprachige, der Rest ist französischsprachig. Beiden gemeinsam sind aber ganze 12 der insgesamt 123

⁹ Vgl. Johan Galtung (1985: 164f. u.ö.) zum Stil der "gallischen" Theoriebildung.

Einträge in Biber et al. (1998) und der insgesamt 233 in Habert et al. (1997). Noch größer als bei Habert et al. (1997) ist übrigens der quantitative Unterschied zwischen englisch- und deutschsprachigen Titeln in der Bibliographie von Lenz (2000). Sie nennt 229 (= 83%) englischsprachige und nur 45 (15%) deutschsprachige Arbeiten. Ganz undurchlässig scheint die deutsch-französische Sprachgrenze in der KL in beiden Richtungen zu sein. Habert et al. (1997) nennen keinen einzigen deutschsprachigen Titel, Lenz (2000) einen einzigen französischsprachigen.

Die Internationalität der KL äußert sich mithin nicht in einer sprachlichen Vielfalt der Forschung und ihrer Dokumentation, sondern im Gebrauch des Englischen als *lingua franca*. Das gilt auch für die Sprache der jeweils benutzen bzw. erwähnten Textkorpora: Biber et al. (1998) listen im Anhang 31 Korpora auf, Meyer (2002) sogar deren 60. Alle bestehen nur aus englischen Texten. Aber auch Habert et al. 1997 nennt auf S. 17f. unter der Überschrift *Principaux corpus cités* 9 englisch-amerikanische und ganze drei französische Korpora. Entsprechend heißt es in der *conclusion* S. 214: "Les ressources pour le français sont encore denrée rare. Il n'existe pas d'équivalent pour le français de **Brown, LOB** et de **BNC**, pour la langue contemporaine, ou d'**Archer**, pour l'histoire de la langue." Der *Trésor de la Langue française TLF*, den Susanne Lenz in der Einführung zu ihrer Bibliographie eigens erwähnt, um festzustellen, dass ein diesem entsprechendes großes allgemeinsprachliches Korpus fürs Deutsche noch fehle, wird von Habert et al. 1997 nur einmal als Material für eine *soft ware* kurz erwähnt.

Fazit: Es hat sich herausgestellt, dass die KL genauso einzelsprachorientiert ist wie etwa die traditionelle Grammatik oder Lexikologie. Die Möglichkeiten der Benutzung von Analysewerkzeugen für mehrere Einzelsprachen dürften sehr beschränkt sein; zumindest ist ihre Adaptation notwendig, um sie für die Beschreibung einer Sprache zu nutzen, für die sie nicht entwickelt wurden. Biber et al. (1998) behandeln ganz selbstverständlich nur englische Korpora und Strukturen, ohne dies im Titel oder Umschlagtext ihres Buches zu signalisieren geschweige denn zu kommentieren, und Habert et al. (1997) bedauern immer wieder das Fehlen von französischen Korpora, von *soft ware*, die deren Analyse erlauben würde, und von Ergebnissen korpuslinguistischer Beschreibungen französischer Strukturen. Ihr ausführliches Referat der GB- und US-englischen KL soll letztlich nur die Desiderate der französischen KL sichtbar machen.

Wie schon oben festgestellt, handelt es sich bei Biber et al. (1998) und Habert et al. (1997) ohne jeden Zweifel um Exemplare zweier verschiedener Textsorten: Biber et al. (1998) haben eine Einführung *sensu strictu* in eine Teildisziplin der Linguistik des Englischen geschrieben, Habert et al. (1997) einen kritischen Forschungsbericht über den *state of the art* der entsprechenden Teildisziplin der Linguistik des Französischen. Interessanter ist die Tatsache, dass es außer Biber et al. (1998) eine Reihe weiterer vergleichbarer englischer Einführungen in die englische KL¹⁰ gibt, während sich die Suche nach auch nur einer einzigen vergleichbaren

¹⁰ Die Einführung von Meyer (2002) kommt Biber et al. (1998) besonders nahe, legt aber weniger Gewicht auf die Motivierung des Laien, dafür bringt sie mehr linguistische Information. McEnery et al. (1996) schließen die

Einführung in die französische (oder deutsche) KL als vergeblich erwiesen hat. Das könnte nun einfach daran liegen, dass die KL des Englischen im universitären Ausbildungssystem der englischsprachigen Länder schon fester etabliert ist als in dem der beiden anderen Sprachräume. Dass eine deutsche Einführung in die KL immerhin schon angekündigt ist, spricht für diese Erklärung. Dazu kommt die *lingua-franca*-Funktion des Englischen. Ob darüber hinaus im Sinne von Johan Galtung (1985) mit Auswirkungen unterschiedlicher Wissenschafts- und Wissenschaftsvermittlungsstile der jeweiligen Forschergemeinschaften zu rechnen ist, muss letztlich offen und der Diskussion in diesem Seminar überlassen bleiben. Vielleicht kennen ja andere Seminarteilnehmer französisch- und/oder deutschsprachige Einführungen in die KL von der Art der Biberschen, so dass wir die vage Idee unterschiedlicher "intellektueller Stile" im oben skizzierten Sinne zumindest für die KL der Gegenwart in der sog. westliche Welt getrost aufgeben können. Und solange man noch gar keine Einführung in die KL hat lesen können, die im "gallischen" oder im "teutonischen Wissenschaftsstil" verfasst wäre, sollte man mit Spekulationen darüber sehr vorsichtig sein, wie die wohl aussehen könnten.

Nachtrag Sommer 2006

Inzwischen – genauer im Frühsommer 2006 – ist die 2004 vergeblich gesuchte deutschsprachige Einführung in die KL des Deutschen erschienen. Unter dem Titel *Korpuslinguistik. Eine Einführung* haben Lothar Lemnitzer und Heike Zinsmeister genau das Buch vorgelegt, dessen damaliges Fehlen den Ausweg wählen ließ, von den Erfahrungen beim Lesen eines englischen und eines französischen Buch mit entsprechendem Titel zu berichten und auf der Basis der damaligen Erfahrungen sogar das sehr problematische Thema der (national)sprachlichen Wissenschaftsstile anzuschneiden. Lemnitzer und Zinsmeister haben für die Zukunft nicht nur jenen Ausweg unnötig gemacht, sie entziehen mit ihrem Buch auch den Spekulationen betr. Wissenschaftsstile, die sich angesichts der tiefgreifenden Unterschiede zwischen den beiden seinerzeit vorgestellten Darstellungen der KL anboten, zumindest partiell die Grundlage.¹¹

Diese positive Gesamtbeurteilung dieses Buches sei im folgenden in einigen Stichworten näher begründet:

Die flott geschriebene¹² Einführung verbindet das Bemühen um Motivation des oftmals direkt angesprochenen Lesers (mit dem sich der Autor dieses Textes durchaus identifizieren kann)

Kapitel regelmäßig mit einer *Chapter summary*, mit *Study questions* (und *Suggested solutions* im Anhang) sowie *Further readings*.

¹¹ Man könnte sich und die Autoren allenfalls fragen, warum sie die konsequente Ausrichtung ihrer Einführung auf die deutsche Sprache nicht auch im Titel ihres Buches anzeigen und damit dem schlechten Vorbild der beiden oben eingehend besprochenen Werke folgen.

¹² So werden die trocken-sachlichen Kapitelüberschriften des Inhaltsverzeichnisses (etwa: *Linguistische Korpora, Linguistische Annotation und ihre Nutzung* oder – das längste Kapitel – *Korpuslinguistik in der Praxis*) im Text durch humorvoll-sprachspielerische Vorspanne eingeführt, und das klingt dann so: *Der Stein der Weisen? Linguistische Korpora, Auf den Schultern anderer stehen – Linguistische Annotation und ihre Nutzung, Wie man in den Wald hineinruft – Korpuslinguistik in der Praxis*.

mit einer kurzen, aber erfreulich ausgewogenen und sachlichen Darstellung der theoretischen Basis (Stichworte: empiristischer versus rationalistischer Ansatz; Induktion versus Deduktion, Performanz- vs. Kompetenzbezug) und mit ausführlichen Informationen über praktische Möglichkeiten und Probleme (zugängliche deutsche Korpora, Suchwerkzeuge, Metadaten, Metazeichen, Reguläre Ausdrücke), über Chancen und Gefahren der KL und wie man letztere vermeiden kann. Der wissenschaftsgeschichtliche Anschluss der KL an den englischen Kontextualismus ist überzeugend und meines Wissens originell.

Besonders wichtig: Das Buch ist zum Selbststudium geeignet. Am Anfang jeden Kapitels erfährt der Leser, was er nach der Lektüre desselben dazugelernt haben wird – man beachte das pädagogisch-optimistische Futur! –, jedem Kapitel folgen sorgfältig ausgewählte weiterführende Literaturempfehlungen und Aufgaben – kein leichtes Unterfangen, das natürlich nicht immer gleich gut gelingt. Die Möglichkeiten einer zusätzlichen begleitenden Webseite werden vielseitig genutzt, unter anderem für Lösungsvorschläge. Schliesslich war es eine gute Idee, Erfahrungen von Linguisten mit Korpora zu erfragen und zusammenfassend darzustellen. Die Beispiele sind stark zweckorientiert ausgewählt: viel Syntax, noch mehr Lexikographie (Anglizismen, Neologismen, Partikeln). Mathematische und statistische Grundlagen werden allenfalls vorsichtig in kleinen Dosen angesprochen.

Was 2004 an der englischen und französischen Darstellung der KL auffiel: das praktische Fehlen wissenschaftlicher Kontakte der französischen mit der deutschen KL, viel Kontakt der französischen mit der anglo-amerikanischen KL, aber weitgehende Selbstgenügsamkeit der anglo-amerikanischen (das alles nach Ausweis der Literaturverzeichnisse), das bestätigt sich an Lemnitzer/Zinsmeister (2006): Von 211 Literaturangaben verweisen 46% auf deutschsprachige Texte, 54% auf englische. Eine einzige gilt einem französischsprachigen Titel und zwar den *Éléments de syntaxe structurale* von Lucien Tesnière, die nur deswegen zitiert werden, weil manche Korpustexte in Anlehnung an dessen Dependenztheorie annotiert werden. Dieses wechselseitige Desinteresse der französischen und deutschen Korpuslinguisten aneinander dürfte auch mit der sprachtypologischen Entfernung der beiden Sprachen voneinander zusammenhängen.

Zum Schluss ein bisschen kleinliche Kritik, die die Qualitäten des Buches in keiner Weise beeinträchtigt:

Der Leser vermisst einige wenige Sacherläuterungen: Was ist "N-Gramm Analyse"? Die in der gleichen Tabelle (S. 33) genannten "Personen" hätten man wohl kurz erläutern müssen. S. 130: Sind Flexive Wortbildungsmittel?, S. 133: Mit *abbaubar* und *verformbar* hat man keine Probleme, wenn man beide auf transitive Verben zurückführt. S. 138: Was ist ein Spezifikator?, S. 151: Sind Bedeutungen Zeichen?

Druckfehler können offensichtlich auch mit Korrekturprogrammen nicht ganz ausgemerzt werden. Eine entsprechende Liste ist den Autoren für ein allfällige Neuauflage zugeschickt worden.

Literaturangaben

- Adamzik, Kirsten (2002): *Kontrastive Textologie*. Tübingen.
- Aijmer, Karin/Altenburg, Bengt (eds.) (1991): *English Corpus Linguistics. Studies in Honor of Jan Svartvik*. London.
- Albert, Ruth/Koster, Cor J. (2002): *Empirie in Linguistik und Sprachlehrforschung. Ein methodologisches Arbeitsbuch*. Tübingen. (= Narr Studienbücher).
- Biber, Douglas/Conrad, Susan/Reppen, Randi (1998): *Corpus Linguistics. Investigating Language Structure und Language Use*. Cambridge: UP. (= Cambridge Approaches to Linguistics).
- Brosius, Hans-Bernd/Koschel, Friedrike (2001): *Methoden der empirischen Kommunikationsforschung. Eine Einführung*. Opladen. (= Studienbücher zur Kommunikations- und Medienwissenschaft).
- Clyne, Michael (1987): "Cultural differences in the organization of academic texts. English and German". *Journal of Pragmatics* 11: 211-247.
- Galtung, Johan (1985): "Struktur, Kultur und intellektueller Stil. Ein vergleichender Essay über sachsenische, teutonische, gallische und nipponische Wissenschaft". In: Wierlacher, Alois (ed.): *Das Fremde und das Eigene. Prolegomena zu einer interkulturellen Germanistik*. München: 151-192.
- Habert, Benoît/Nazarenko, Adeline/Salem, André (1997): *Les linguistiques de corpus*. Paris.
- Habert, Benoît/Fabre, Cécile/Issac, Fabrice (1998): *De l'écrit au numérique. Constituer, normaliser et exploiter les corpus électroniques*. Paris.
- Kaplan, Robert B. (1966): "Cultural thought patterns in inter-cultural education". *Language Learning* 16: 1-20.
- Kennedy, Graeme (1998): *An Introduction to Corpus Linguistics*. London.
- Lenders, Winfried/Willée, Gerd (1998): *Linguistische Datenverarbeitung*. 2., neu bearbeitete und erweiterte Aufl. Opladen.
- Lemnitzer, Lothar/Zinsmeister, Heike (2006): *Korpuslinguistik. Eine Einführung*. Tübingen. (= Narr Studienbücher).
- Lenz, Susanne (2000): *Korpuslinguistik*. Tübingen. (= Studienbibliographien Sprachwissenschaft 32).
- McEnery, Tony/Wilson, Andrew (2001): *Corpus Linguistics. An Introduction*. Edinburgh. (= Edinburgh textbooks in empirical linguistics).
- Meyer, Charles (2002): *English Corpus Linguistics. An Introduction*. Cambridge.
- Pusch, Claus D./Raible, Wolfgang (eds.) (2002): *Romanistische Korpuslinguistik. Romance Corpus Linguistics. Korpora und gesprochene Sprache. Corpora and Spoken Language*. Tübingen. (= ScriptOra 126).
- Schlobinski, Peter (1996): *Empirische Sprachwissenschaft*. Opladen. (= WV Studium 174).

Anhang 1:**Die 38 Beiträge zu Pusch et al. (2002):**

Objekt(sprache), Sprache der Redaktion und sprachregionale Zugehörigkeit des Autors

Objekt(sprache)	Sprache der Redaktion	sprachregionale Herkunft des Autors	Anzahl
Französisch	Französisch	Frankophonie	9
Romanisch	Französisch	Frankophonie	1
Französisch	Französisch	Italien	1
Französisch	Französisch	Anglophonie	1
Französisch	Französisch	Germanophonie	2
Romanisch	Französisch	Germanophonie	1
Französisch	Französisch	Dänemark	1
Katalanisch	Französisch (als LF)	Spanien	2
Korpuslinguistik	Französisch (als LF)	Spanien	1
Spanisch	Französisch (als LF)	Germanophonie	1
Korpuslinguistik	Deutsch	Germanophonie	2
Romanisch	Deutsch	Germanophonie	2
Deutsch+Franz.	Deutsch	Germanophone	2
Letzebuergisch	Deutsch	Luxemburg	1
Portugisich	Deutsch	Germanophonie	1
Italienisch	Deutsch	Germanophonie	1
Romanisch	Englisch (als LF)	Germanophonie	2
Italienisch	Englisch (als LF)	Germanophonie	1
Katalanisch	Englisch (als LF)	Spanien	1
Portugisich	Englisch (als LF)	Brasilien	1
Spanisch	Englisch	Anglophonie	1
Französisch	Englisch	Anglophonie	1

Es fehlen die Kombinationen von Französisch als Objektsprache und Englisch oder Deutsch als Sprache der Redaktion.

Der Gebrauch einer Sprache für die Redaktion gilt als ein *lingua franca*-Gebrauch (= LF) dieser Sprache, wenn ihre Wahl weder durch den Gegenstand des Beitrags noch durch die sprachregionale Zugehörigkeit des Autors motiviert sein kann. Die 4 französischsprachigen und 5 englischsprachigen Beiträge, für die die betreffende Sprache damit als *lingua franca* funktioniert, sind in der 2. Spalte durch den Zusatz "als LF" markiert. Angesichts der kleinen Anzahlen kann dieses Ergebnis bestenfalls eine Tendenz in der gegenwärtigen romanistischen KL anzeigen.

Anhang 2:

Synoptische Darstellung der Struktur von Biber et al. (1998) und Habert et al. (1997)

Biber et al. (1998) [300 S.]	Habert et al. (1997) [240 S.]
<p>1. Introduction: Goals and methods of the corpus-based approach [18 S.] 1.1. Studying language: structure and use 1.2. What is the corpus-based approach? 1.3. Corpora and corpus analysis tools used in this book [4 engl.Korpora] 1.4 Overview of the book</p> <p>I. Investigating the use of language features 2. Lexicography [33 S.] 3. Grammar [29 S.] 4. Lexico-grammar [21 S.] 5. The study of discourse characteristics [24 S.]</p> <p>II. Investigating the characteristics of varieties 6. Register variation and English for Specific Purposes [35 S.] 7. Language acquisition and development [31 S.] 8. Historical and stylistic investigations [25 S.]</p> <p>III. Summing up and looking ahead 9. Conclusion [8 S.]</p> <p>IV. Methodological boxes [35 S.] 1. Issues in corpus design 2. Issues in diachronic corpus design 3. Concordancing packages vs. programming for corpus analysis 4. Characteristics of tagged corpora 5. The process of tagging 6. Norming frequency counts 7. Statistical measures of lexical associations 8. The unit of analysis in corpus based studies 9. Significance tests and the reporting of statistics 10. Factor loading and dimension scores</p>	<p>Introduction [12 S.] 1. Le regain d'intérêt pour les corpus 2. A quoi servent les corpus annotés? 3. Choix terminologiques 4. Notations 5. Orientation de l'ouvrage 6. Démarche suivie 7. Principaux corpus cités [10 engl.,3 frz.]</p> <p>I. Les corpus annotés et leurs utilisations 1. Les corpus étiquetés [22 S.] 2. Les corpus arborés [30 S.] 3. Les ressources lexicales pour l'étiquetage sémantique [19 S.]</p> <p>II. Dimensions transversales 4. Des mots aux sens: sémantique en corpus [25 S.] 5. Le langage au fil du temps: corpus et diachronie [14 S.] 6. D'une langue à l'autre: les corpus alignés [6 S.]</p> <p>III. Méthodes et techniques 7. Constituer un corpus [17 S.] 8. Annoter un corpus [22 S.] 9. Quantifier les faits langagiers [29 S.]</p> <p>Conclusion [5 S.]</p>

Appendix: commercially available corpora [33 engl.] and analytical tools [11 engl.] References [123 ausschl. engl. Einträge] Index [3 1/2 S., nur Termini]	Bibliographie [233 Einträge; engl./frz: 2/1] Index [14 Sp., Termini und Namen]
---	---