

Digitale Sprachressourcen in Lehramtsstudiengängen: Kompetenzen – Erfahrungen – Desiderate

PREPRINT. Erscheint in: Frank Binder, Henning Lobin & Harald Längen (Hrsg., 2011): Sprachressourcen in der Lehre (Journal for Language Technology and Computational Linguistics, Themenheft).

- 1 Einleitung
- 2 Motivation und curriculare Einbettung
 - 2.1 Generelle Überlegungen
 - 2.2 Anknüpfungspunkte an schulische Lehrpläne
 - 2.3 Anknüpfungspunkte an universitäre Lehramtsstudiengänge
- 3 Seminar „Korpusgestützte Sprachanalyse“
 - 3.1 Gegenstand, Szenario, genutzte Ressourcen
 - 3.2 Erfahrungen mit Fehlerquellen und Grenzen der aktuellen Korpus-technologie
- 4 Einsatz digitaler Sprachressourcen in Lehrveranstaltungen zum Thema „Internetbasierte Kommunikation“
 - 4.1 Gegenstand und Seminarszenario
 - 4.2 Genutzte Ressourcen / Dortmunder Chat-Korpus
 - 4.3 Erfahrungen und Herausforderungen
- 5 Desiderate und Perspektiven

1 Einleitung

In diesem Artikel beschreiben wir,

- wie sich die Arbeit mit digitalen Sprachressourcen in die Curricula der Hochschulgermanistik und der Lehrpläne für das Fach Deutsch an Schulen integrieren lässt und weshalb wir die Vermittlung entsprechender Kompetenzen in Lehramtsstudiengängen für wichtig und hochgradig berufsfeldrelevant halten,
- wie wir in der Dortmunder Linguistik und Sprachdidaktik Sprachressourcen in der Lehre einsetzen (und welche),
- welche Erfahrungen wir dabei gemacht haben und welche Wünsche und Anregungen zur Erleichterung des didaktischen Einsatzes von Sprachressourcen sich daraus ableiten lassen.

Die allgemeinen Überlegungen zur Relevanz des Themas für die Lehrerbildung und zur curricularen Einbettung in entsprechende Studiengänge werden ergänzt um die Beschreibung zweier Seminarkonzepte zu den Themen „Korpusgestützte Sprachanalyse“ und „Internetbasierte Kommunikation“ (Abschnitte 3 und 4). Im einen Fall ist die Vermittlung von Methodenkompetenz in Bezug auf die Nutzung digitaler Sprachressourcen selbst Gegenstand der Veranstaltung, im anderen Fall werden Sprachressourcen punktuell als Hilfsmittel für die Durchführung kleiner Analyseprojekte genutzt. Beide Seminarkonzepte wurden bereits mehrfach erprobt und dabei z.T. auch weiterentwickelt. In Abschnitt 5 formulieren wir auf der Grundlage unserer bisher gemachten Erfahrungen Desiderate für die Ausgestaltung digitaler Sprachressourcen in diesem Bereich.

Sämtliche im Artikel mit Kurztiteln erwähnten Sprachressourcen sind in Abschnitt 6.2 mit ihren vollständigen Bezeichnungen und zugehörigen URL-Verweisen aufgeführt.

2 Motivation und curriculare Einbettung

2.1 Generelle Überlegungen

Zwei generelle Überzeugungen motivieren den Einsatz von digitalen Sprachressourcen in der Ausbildung von Lehramtsstudierenden am Institut für deutsche Sprache und Literatur der TU Dortmund:

(1) Die fachwissenschaftliche Ausbildung in deutscher Linguistik sollte nicht nur Theorien, Beschreibungskategorien und -modelle vermitteln, sondern auch die Methoden, die der Modell- und Kategorienbildung zugrunde liegen. Erst dies befähigt angehende LehrerInnen dazu, didaktische Konzepte des „entdeckenden Lernens“, wie sie z.B. in der „Grammatikwerkstatt“ (EISENBERG & MENZEL 1999) oder im funktionalen Grammatikunterricht (HOFFMANN 2006) angestrebt werden, im schulischen Sprachunterricht souverän umzusetzen. Da digitale Sprachressourcen, insbesondere digitale Text- und Gesprächskorpora, für die empirisch geleitete linguistische Kategorien- und Modellbildung an Bedeutung gewinnen, werden entsprechende Methoden auch für Lehramtsstudierende relevant, selbst wenn diese an eigener linguistischer Forschung nicht primär interessiert sind. Weiterhin sind digitale Sprachressourcen im Internet natürlich auch eine Fundgrube für Unterrichtsmaterial im Rahmen von internetbasierten Didaktikkonzepten wie z.B. WebQuests.

(2) Die Kompetenz zur kritischen Bewertung und aufgeklärten Nutzung digitaler Sprachressourcen zur deutschen Sprache (lexikalische und enzyklopädische Ressourcen, Korpora) wird für angehende DeutschlehrerInnen zunehmend relevant. Gerade weil viele verschiedene Typen von Ressourcen im Internet verfügbar sind und sich als potenzielle Datenquellen zur deutschen Sprache anbieten (vgl. ENGELBERG & LEMNITZER 2009; STORRER 2010), ist es wichtig, Schülerinnen und Schülern Leitlinien für die Bewertung der Qualität und der Zuverlässigkeit dieser Daten zu vermitteln. Dies gilt für kollaborativ aufgebaute lexikalische Ressourcen, z.B. *Wiktionary*, das mehrsprachige *Dict.cc*, den

OpenThesaurus, ebenso wie für retrospektiv digitalisierte Wörterbücher, z.B. *DWDS-WDG* oder *DWB-Online*, deren Quellenwert an den ursprünglichen Entstehungskontext angebunden werden muss. Auch für die aufgeklärte und kundige Nutzung digitaler Sprachkorpora müssen Kriterien vermittelt werden, anhand derer Studierende bewerten können, welche Aussagekraft die bei der Korpusrecherche erzielten Ergebnisse für eine bestimmte Fragestellung haben.

Wichtig ist uns, dass wir die Nutzung von digitalen Sprachressourcen nicht als eine linguistische „Schule“ einführen, sondern als einen methodischen Zugang zu Sprache, der parallel und ggf. auch in Kombination mit anderen methodischen Zugängen dazu beiträgt, unser Wissen über sprachliche Strukturen und Funktionen zu erweitern. Gerade damit der Stellenwert datengestützter Zugänge im methodischen Spektrum der Sprachwissenschaften besser bestimmt werden kann, muss nach unserer Erfahrung nicht nur das computertechnische Rüstzeug – z.B. die Syntax von Abfragesprachen – vermittelt werden, sondern auch die Kompetenz zur kritischen Bewertung der Qualität von Daten und der Aussagekraft von datengestützten Analysen. Wir ermuntern die Studierenden dazu, sich vor der Nutzung von digitalen Ressourcen Fragen wie die folgenden zu beantworten:

- Wer baut die Ressource auf / bearbeitet sie redaktionell / bietet sie an?
- Welche Maßnahmen der Qualitätskontrolle werden von Seiten der Anbieter thematisiert? Wie kann man als Nutzer die Qualität der Daten evaluieren? Welche qualitativen Aspekte sind für meine Fragestellung besonders relevant.
- Welchen Sprachausschnitt beschreibt/dokumentiert die Ressource? Wie umfangreich ist sie (im Vergleich zu Ressourcen, die ggf. als Alternativen zur Verfügung stehen)?
- Eignet sich die Ressource überhaupt für die Bearbeitung meiner Fragestellungen; welche Einschränkungen gibt es?
- Wie ist die Ressource strukturiert? Welche Typen von Daten sind systematisch mit welchen Beschreibungskategorien erfasst?

Der Umgang mit digitalen Textkorpora stellt dabei besonders hohe Ansprüche: Für die vergleichende Bewertung von Korpusressourcen müssen grundlegende Konzepte der Korpuslinguistik (Annotation, Metadaten, Lemmatisierung, Wortartentagging etc.) und der Qualitätsbewertung verstanden worden sein. Wie wir in Abschnitt 3 noch weiter ausführen werden, ist es nach unserer Erfahrung sinnvoll, die Vermittlung dieser Konzepte mit kleinen praktischen Aufgaben und Beispielen zu verknüpfen und auf der Basis der dabei gemachten Erfahrungen Qualitätsaspekte und methodische Herausforderungen zu reflektieren.

Wie und mit welchen Kompetenzzielen digitale Ressourcen im Unterricht in verschiedenen Schulstufen nutzbar sind, muss noch weiter erprobt werden. Nach unserer Erfahrung sind Studierende sehr gut motivierbar, didaktische Konzepte zu entwickeln und in schulischen Praxisphasen zu erproben; sie stoßen dabei erfahrungsgemäß auf großes Interesse bei den Schülerinnen und Schülern aller Altersstufen und meist auch auf Interesse

seitens der Schulleitungen und der betreuenden Lehrerinnen und Lehrer. Die Praxisphasen in den Lehramtsstudiengängen bieten hierfür Anknüpfungspunkte für Experimente.

2.2 Anknüpfungspunkte an schulische Lehrpläne

Bislang ist der Einsatz digitaler Sprachressourcen nicht explizit in den Kompetenzprofilen schulischer Lehrpläne verankert. Die verschiedenen Formen der Ressourcennutzung und die dafür zu vermittelnden Kompetenzen sind aber problemlos an übergreifende Kompetenzentwicklungsziele anschließbar. Im Folgenden seien hierfür nur einige Beispiele aus den NRW-Lehrplänen genannt:

(1) *Wissenschaftspropädeutik*: Als eine der beiden zentralen Aufgaben der gymnasialen Oberstufe nennen die Richtlinien und Lehrpläne des Landes Nordrhein-Westfalen die *Wissenschaftspropädeutik* – im Sinne eines „wissenschaftsorientierte[n] Lernen[s], das durch Systematisierung, Methodenbewusstsein, Problematisierung und Distanz gekennzeichnet ist“ (RLP Sek II GyGe Deutsch NRW: XII). Ein wichtiges Vermittlungsziel im Rahmen des wissenschaftspropädeutischen Lernens besteht darin, die Schülerinnen und Schüler dazu zu befähigen, „grundlegende wissenschaftliche Erkenntnis- und Verfahrensweisen systematisch [zu] erarbeiten“ und dabei zu lernen, „Aufgabenstellung[en] selbstständig zu strukturieren, die erforderlichen Arbeitsmethoden problemangemessen und zeitökonomisch auszuführen, Hypothesen zu bilden und zu prüfen und die Arbeitsergebnisse angemessen darzustellen“ (ebd.). Die Verknüpfung unterrichtlicher Sprachreflexion mit einer Analyse authentischer Sprachdaten (Korpora) kann in diesem Zusammenhang einerseits Zugänge zur Arbeitsweise einer modernen datengestützten Linguistik eröffnen und andererseits einen Ausgangspunkt für die Vermittlung allgemeiner wissenschaftspropädeutischer Grundlagen zum Verhältnis von wissenschaftlicher Theoriebildung und Empirie im Allgemeinen bilden. Da die Wissenschaftspropädeutik eine Aufgabe nicht nur des Deutsch-, sondern jedweden Fachunterrichts in der gymnasialen Oberstufe darstellt, sind hier auch fächerübergreifende Konzepte denkbar.

(2) *Methodenkompetenz*: In der didaktischen Konzeption des Faches Deutsch wird neben der sprachlichen, kulturellen, ästhetischen und ethischen Kompetenz die *Methodenkompetenz* als eines von fünf zentralen Vermittlungszielen herausgestellt. Der Fokus liegt dabei auf „Methoden sprachlichen Arbeitens (Methoden des Verstehens, Methoden sprachanalytischer Arbeit, Methoden schriftlicher Darstellung, Methoden mündlicher Verständigung, Methoden produktionsorientierten Arbeitens), die gleichzeitig Unterrichtsgegenstand, fachliche Verfahrensweisen und Lernstrategien darstellen“ (RLP Sek II GyGe Deutsch NRW: 6). In allen Teilaspekten liegt die Arbeit mit digitalen Sprachressourcen nahe, z.B.:

- Als Methoden für das Verstehen und die Erschließung von Texten bilden Kenntnisse im Umgang mit Nachschlagewerken und zu den in Nachschlagewerken unterschiedlichen Typs verfügbaren Informationstypen eine unerlässliche Grundlage. Die „Beschaf-

fung und Bearbeitung von Informationen“ und die „Nutzung von Hilfsmitteln, auch in elektronischen Netzen“ (ebd.: 27) werden in diesem Zusammenhang als zentrale Vermittlungsziele im Bereich „Methoden des Textverstehens“ benannt – ein Bereich, in den sich die Arbeit mit digitalen Sprachressourcen (z.B. digitalen Allgemeinwörterbüchern zur deutschen Gegenwartssprache, online verfügbaren Autoren- und historischen Wörterbüchern) unmittelbar integrieren lässt.

- Zu den „Methoden schriftlicher Darstellung“ zählt sicher auch das Nachschlagen in Online-Nachschlagewerken zur deutschen Sprache; hierfür müssen Kompetenzen zur Bewertung von Qualität und Zuverlässigkeit vermittelt werden. Eine „effektive Nutzung der fachspezifischen Informations- und Kommunikationsangebote in Bibliotheken und elektronischen Netzen“ (ebd.: 28) ist hier explizit als ein Kompetenzziel benannt, gerade auch im Zusammenhang mit der wissenschafts-propädeutischen Aufgabe des Deutschunterrichts.
- Für die Vermittlung von „Methoden sprachanalytischer Arbeit“ bietet sich, insbesondere in Kombination mit dem Bereich „Reflexion über Sprache“, die Arbeit mit digitalen Korpora an. Mit Blick auf die Aufgabe des sprachbezogenen Deutschunterrichts, die Schülerinnen und Schüler zu einem „norm- und regel-bewussten Sprechen und Schreiben“ und zu einer funktional angemessenen Differenzierung von Sprachverwendungsweisen und stilistischen Varianten zu befähigen (ebd.: 22), kann die Arbeit mit Korpora eine wichtige Bereicherung des didaktischen Instrumentariums darstellen, weil sie bei entsprechender Einführung und der Formulierung geeigneter Arbeitsaufgaben dazu genutzt werden kann, die Schülerinnen und Schüler auf die Besonderheiten stilistischer Variation in unterschiedlichen Zweckbereichen und situativen Kontexten sprachlichen Handelns selbst aufmerksam werden zu lassen.

Um den Umgang mit digitalen Sprachressourcen in der Schule vermitteln zu können, müssen Lehrerinnen und Lehrer selbst reflektierte Kenntnisse zu entsprechenden Ressourcen besitzen und über Möglichkeiten und Rahmenbedingungen orientiert sein, diese didaktisch sinnvoll im Unterricht einzusetzen. Im folgenden Abschnitt möchten wir deshalb am Beispiel der Dortmunder Lehramtsstudiengänge skizzieren, wie sich solche Kompetenzen in die Curricula integrieren lassen.

2.3 Anknüpfungspunkte an universitäre Lehramtsstudiengänge

Zu den übergreifenden Zielen der Dortmunder Lehramtsstudiengänge im Fach Deutsch zählt die Vermittlung von „Erfahrungen in der selbständigen wissenschaftlichen Arbeit“ (FSB Germ BfP Dortmund: 39). Der Umgang mit digitalen Sprachressourcen kann hierzu in mehrerlei Hinsicht beitragen: Digitale Ressourcen wie Korpora und Informationssysteme zur Lexik und Grammatik können für die eigenständige wissenschaftliche Arbeit genutzt werden; hierbei werden deren Potenziale und Besonderheiten im Rahmen konkreter Nutzungsszenarien (z.B. Nachschlageaufgabe, Analyseprojekt) deutlich – eine

wichtige Voraussetzung, um einschätzen zu können, ob und wie diese Ressourcen auch im schulischen Deutschunterricht einsetzbar sind. In den fachdidaktischen Anteilen des Studiums kann der Einsatz digitaler Sprachressourcen in der Schule explizit thematisiert und im Rahmen von Praxisphasen erprobt werden. Dieser curriculare Aufbau – Einführung in den Umgang mit Sprachressourcen → Anwendung in fachwissenschaftlichen Analyseprojekten → Entwicklung fachdidaktischer Konzepte für den schulischen Einsatz → Erprobung und Evaluation in der Schulpraxis – ist prinzipiell in alle Phasen und Aspekte des Lehramtsstudiums (die einführenden wie die aufbauenden, die fachwissenschaftlichen wie fachdidaktischen) integrierbar.

In den Lehramtsstudiengängen an der TU Dortmund werden bereits in den sprachwissenschaftlichen Pflichtveranstaltungen (eine 4-stündige „Einführung in die Sprachwissenschaft“ und ein 2-stündiges Proseminar „Grundlagen der Grammatik“) digitale Sprachressourcen – insbesondere *Canoo.net*, *ProGr@mm* und *Grammis* – für kleine Recherche- und Analyseaufgaben genutzt; in diesem Zusammenhang werden auch Aspekte der Qualität und Verlässlichkeit von digitalen Quellen thematisiert.

In weiterführenden Wahlpflichtveranstaltungen werden diese Grundkenntnisse vertieft und um weitere Ressourcentypen – z.B. digitale Korpora und lexikalische Informationssysteme – erweitert. Dabei lassen sich zwei Typen der Integration digitaler Sprachressourcen in die didaktischen Konzepte der Lehrveranstaltungen unterscheiden:

Typ 1: Seminare, in denen digitale Sprachressourcen den *Gegenstand* der Veranstaltung bilden (z.B. zu den Themen „Digitale Nachschlagewerke zur deutschen Sprache“, „Korpusgestützte Sprachanalyse“, „Internetbasierte Lexikographie“). In Veranstaltungen dieses Typs werden – je nach Seminarthema – verschiedene Kombinationen von digitalen Korpora, lexikalischen und grammatischen Informationssystemen und digitalen Wörterbüchern genutzt. Lehr-/Lernziele sind die Vermittlung von Methodenwissen zum Umgang mit Sprachressourcen, die Erarbeitung von Qualitätskriterien und die kritische Reflexion ihres Nutzungspotenzials in verschiedenen Anwendungsfeldern (insbesondere Sprachdidaktik und Lexikographie). In Abschnitt 3 wird ein Seminarkonzept dieses Typs im Detail vorgestellt.

Typ 2: Seminare, in denen digitale Sprachressourcen als *Hilfsmittel und empirische Basis* für Analyseprojekte der Studierenden genutzt werden (z.B. zu den Themen „Wortbildung und Wortschatzentwicklung“, „Phraseologie des Deutschen“, „Orthographie des Deutschen“, „Mediale Bedingungen des kommunikativen Handelns“, „Internetbasierte Kommunikation“). Die Art der Analyseaufgaben und die dafür geeigneten Ressourcen ergeben sich aus dem jeweiligen Seminarthema. Zum Einsatz kamen bislang z.B. die Korpora und Wörterbücher im *DWDS*-Portal, die *Canoo*-Spezialwörterbücher zur Wortbildung und zur Orthographie, die *Wortwarte* und das *Szenesprachenwiki* zur Wortschatzentwicklung im Gegenwartsdeutschen, das Neologismen- und das phraseologische Wörterbuch in der *eLexiko*-Komponente von *OWID* sowie Spezialkorpora zur internetbasierten Kommunikation. In Abschnitt 4 werden wir ein Seminarkonzept dieses Typs zum

Thema Internetbasierte Kommunikation vorstellen, das mit am Lehrstuhl aufgebauten Korpora zur Chat-Kommunikation arbeitet.

3 Seminar „Korpusgestützte Sprachanalyse“

3.1 Gegenstand, Szenario, genutzte Ressourcen

Für die kundige Nutzung digitaler Textkorpora benötigt man sowohl korpuslinguistisches Grundwissen als auch Leitlinien für die methodische Planung und Durchführung korpusgestützter Analysen. An der TU Dortmund haben wir ein Seminarkonzept erprobt und weiterentwickelt, um diese Kombination von korpuslinguistischen und empirisch-methodischen Kompetenzen aufzubauen. Die Seminare sind verankert im Wahlpflichtbereich des Moduls „Anwendungsfelder der Sprach-, Literatur- und Medienwissenschaft“ der Dortmunder Lehramtsstudiengänge. Das Seminarkonzept umfasst einen Teil, in dem die methodischen Grundlagen der empirischen Sprachanalyse vermittelt werden, z.B. die Unterscheidung von hypothesenerkundendem und hypothesenprüfendem Vorgehen, Leitlinien zur Formulierung falsifizierbarer Hypothesen und zur Bewertung der Validität und Reliabilität von Ergebnissen, die Frage der Repräsentativität von Stichproben, der Stellenwert von Signifikanztests etc. Als Basisliteratur dienen BORTZ & DÖRING (2006), für den Umgang mit Excel zusätzlich ALBERT & KOSTER (2002).¹ Ein weiterer Bestandteil des Seminarkonzepts ist ein Überblick über digitale Sprachressourcen mit dem Fokus auf deutschen Korpora (vgl. die Auswahl in STORRER 2011: Kap.3). Im Anschluss an diesen Überblick werden Grundbegriffe der Korpuslinguistik eingeführt, z.B. die Unterscheidung von Primär- und Metadaten, Formen der linguistischen Aufbereitung wie Lemmatisierung oder POS-Tagging, Verfahren der automatischen Kollokationsanalyse, die Qualitätsbewertung von Suchanfragen durch Genauigkeit (*precision*) und Ausbeute (*recall*) etc. Basisliteratur für diesen Teil bildet die Einführung von LEMNITZER & ZINSMEISTER (2006), ergänzt um MCENERY et al. (2006). Erstmals wurde das Seminar im WS 2006/07 angeboten. Dabei handelte es sich um eine wöchentlich stattfindende Veranstaltung, in der zunächst die Vermittlung der o.g. Konzepte und Methoden am Beispiel von Korpusarbeiten anderer ForscherInnen im Vordergrund stand. Erst gegen Ende des Seminars konzipierten die Studierenden eigene Fragestellungen, die sie nach Seminarende mit Hilfe von Korpusanalysen im Rahmen von Hausarbeiten weiterverfolgten. Der dabei entstandene Betreuungsbedarf sowie entsprechende Rückmeldungen zeigten, dass den Studierenden die Relevanz der im Seminar besprochenen korpuslinguistischen Grundbegriffe und der Nutzwert der linguistischen Aufbereitung (Lemmatisierung, POS-Tagging) erst später beim eigenen Arbeiten deutlich geworden waren. Insbesondere die Möglichkeiten, durch geschickte Abfragestrategien die Ergebnisse von Korpusanalysen zu optimieren,

1 Wie beziehen uns hier auf die Auflagen, die bei der ersten Durchführung im Wintersemester 2006/07 verfügbar waren; in späteren Seminaren wurden jeweils die aktuellsten Auflagen genutzt.

wurden erst im Zuge der eigenen Projekte als sinnvoll und wichtig entdeckt (obwohl sie im Seminar an Beispielen behandelt worden waren).

1) In Bodo Mrotzeks „Lexikon der bedrohten Wörter“ sind folgende Wörter verzeichnet: *Dusel, Pappenheimer, Schabernack, Schelm*.
Bestätigen die Daten des DWDS-Kernkorpus und des DWDS-Zeitungskorpus „Die ZEIT“ die Annahme, dass diese Wörter vom „Aussterben“ bedroht sind?

2) Versuchen Sie, für das Verb „eintrudeln“ möglichst viele Belege zu finden, d.h. auch Belege, in denen die Partikel mit dem Verb die Satzklammer bildet (...trudelt ... ein).

3) Im Zuge der Orthographiereform wurde der Bereich der Getrennt- und Zusammenschreibung kontrovers diskutiert. Untersuchen Sie im DWDS-Kernkorpus, welche der Schreibalternativen von „Eis laufen“ vs. „eislaufen“ in welchen Zeitabschnitten des 20. Jahrhunderts häufiger belegt ist.

Notieren Sie bitte bei der Lösung der Aufgaben 1–3 nicht nur die Ergebnisse, sondern auch die Abfragen, die Sie gemacht haben und ggf. Auffälligkeiten in den dabei erzielten Trefferlisten.

4) Es gilt als ein Trend der (gesprochenen) Gegenwartssprache, dass „weil“ nicht mehr nur als unterordnende Konjunktion (mit Verbendstellung: ... *weil das Wetter schlecht ist*), sondern auch als nebenordnende Konjunktion (mit Verbzweitstellung ... *weil das Wetter ist schlecht*) verwendet wird.

Mit der Suchanfrage "**weil \, "**" kann man im Korpus gesprochener Sprache des DWDS die Suche nach „falschen“ weil-Sätzen auf 117 eingrenzen.

- Welche Arten von Pseudotreffern gibt es?
- Inwiefern könnten die Daten belegen, dass es sich tatsächlich um eine neuere Entwicklung handelt?
- Kann man die Präzision der Abfrage noch weiter optimieren?
- Mit welchen Strategien kann man die Vollständigkeit der Abfrage noch weiter erhöhen?

Abbildung 1: Beispiele für Rechercheaufgaben in den Praxisteilen des Seminars „Korpusgestützte Sprachanalyse“

Aus dieser Erfahrung heraus wurde das Konzept in den Folgeseminaren (im SS 2008 und im SS 2009) so umgestaltet, dass die Vermittlung methodischer und konzeptioneller Grundlagen jeweils verzahnt war mit konkreten kleinen Analyseaufgaben, an denen die Relevanz der Konzepte und Lösungsstrategien für bestimmte Problemstellungen deutlich wird. Dabei erwies sich die Organisationsform als einwöchiges Blockseminar als günstig, das in einem Seminarraum mit angeschlossenem Rechnerraum (mit 30 Plätzen) durchgeführt wurde. Die in den Vormittagssitzungen vermittelten Methoden, Konzepte und Abfragestrategien wurden von Arbeitsgruppen nachmittags am Beispiel kleiner Analyseaufgaben erprobt, dabei standen AnsprechpartnerInnen (Seminarleiterin/TutorInnen) für Rückfragen direkt zur Verfügung. Die Erfahrungen und Ergebnisse der Arbeitsgruppen wurden wiederum im Plenum vorgestellt und diskutiert; diese Diskussionen bildeten gute Anknüpfungspunkte für die zunächst anwendungsunabhängig eingeführten Konzepte und Methoden. Die Studierenden waren für die praktischen Arbeiten sehr gut zu motivieren und entwickelten viel Kreativität beim Experimentieren und Optimieren von Abfragen. Um Unterschiede im Tempo der Arbeitsgruppen aufzufangen, gab es in den AG-Phasen immer mehrere Aufgaben, die in frei wählbarer Reihenfolge, aber nicht unbedingt vollständig zu bearbeiten waren. Um zu vermeiden, dass sich die Studierenden in zu viele unterschiedliche Abfragesprachen eindenken mussten, bezogen sich die meisten Aufgaben auf die Online-Schnittstelle zu den DWDS-Korpora (GEYKEN 2007). Abb. 1 zeigt einige Beispiele für solche Aufgaben; weitere Beispiele und Lösungskommentare dazu – mit

Fokus auf dem Anwendungsbereich Lexikographie – finden sich in den Übungen zu STORRER (2011). Am Ende des Blockseminars wurden Projektpläne für korpusbasierte Studien entworfen und im Plenum diskutiert. Als Leitfaden für die Projektplanung diente das Schema in Abb. 2, das vorab im Seminar an Beispielen aus eigenen korpusgestützten Arbeiten bzw. von der Seminarleiterin betreuten Staatsexamens- und Masterarbeiten vorgestellt und konkretisiert worden war. Die Studierenden konnten entweder vorgegebene Fragestellungen bearbeiten oder Pläne für eigene Projekte entwerfen. Die meisten Studierenden entschieden sich für Letzteres und einige realisierten die geplanten Studien im Rahmen von Hausarbeiten (sofern eine Hausarbeit für die von den Studierenden gewünschte Kreditierung erforderlich war). Der Betreuungsaufwand bei diesen Hausarbeiten erwies sich als erheblich geringer als beim ersten, weniger praxisorientierten Seminar-durchlauf, da durch die Praxisphasen die Kompetenz zur Formulierung geeigneter Abfragen schon vorhanden war und der Aufwand für die intellektuelle Nachanalyse besser abgeschätzt werden konnte.

Die Fragestellungen, die in den Praxisteilen bearbeitet werden, sind wegen der begrenzten Zeit, die für die Analysen zur Verfügung steht, natürlich vergleichsweise simpel. Deshalb wurde das Spektrum der Möglichkeiten korpusunterstützten und korpusbasierten Arbeitens erweitert durch Hinweise auf komplexere Studien in der weiterführenden Literatur, z.B. auf die die Spezialartikel der HSK-Bände zur Korpuslinguistik (LÜDELING & KYTÖ 2008/2009) und zu korpusbasierten Ansätzen in speziellen linguistischen Phänomen- und Anwendungsbereichen (z.B. Phraseologie, Wortbildung; Lexikographie, Grammatikographie). Allerdings sprengen die meisten publizierten Studien den zeitlichen Rahmen, der für studentische Arbeiten und erst recht später für kleine Korpusstudien im schulischen Deutschunterricht zur Verfügung steht. Für das Anliegen, das Arbeiten mit Korpora auch für die Lehrerbildung attraktiv zu machen, ist es deshalb hilfreich, Studierende über kleine Übungen an die Chancen, aber auch an die Grenzen der aktuellen Korpus-technologie heranzuführen. Gerade computerlinguistisch nicht ausgebildete Korpusnutzer überschätzen nämlich tendenziell die Möglichkeiten der automatischen Korpusanalyse und unterschätzen den manuell-intellektuellen Aufwand, der für die Beantwortung vieler linguistischer Fragestellungen immer noch betrieben werden muss. Dies liegt daran, dass sie die automatischen Verfahren zur (computer-)linguistischen Aufbereitung von Korpora nicht durchschauen und deshalb nicht einschätzen können, in welchen Bereichen beim aktuellen Stand der Technologie noch mit vielen Fehlern gerechnet werden muss und welche (prinzipiell interessanten) Fragen ggf. auch überhaupt noch nicht mit vertretbarem Aufwand in digitalen Korpora untersucht werden können. Wenn die Kompetenzen zur Auswahl geeigneter Fragestellungen und geeigneter Heuristiken für die Datenerhebung und -analyse fehlen, kann das Arbeiten mit Korpora schnell als mühselige „Erbsenzähl-Linguistik“ erscheinen. Es ist deshalb ein wichtiges Anliegen des Seminarkonzepts, durch die kleinen Analyseaufgaben typische Fehlerquellen und Grenzen der aktuellen Korpus-technologie deutlich zu machen.

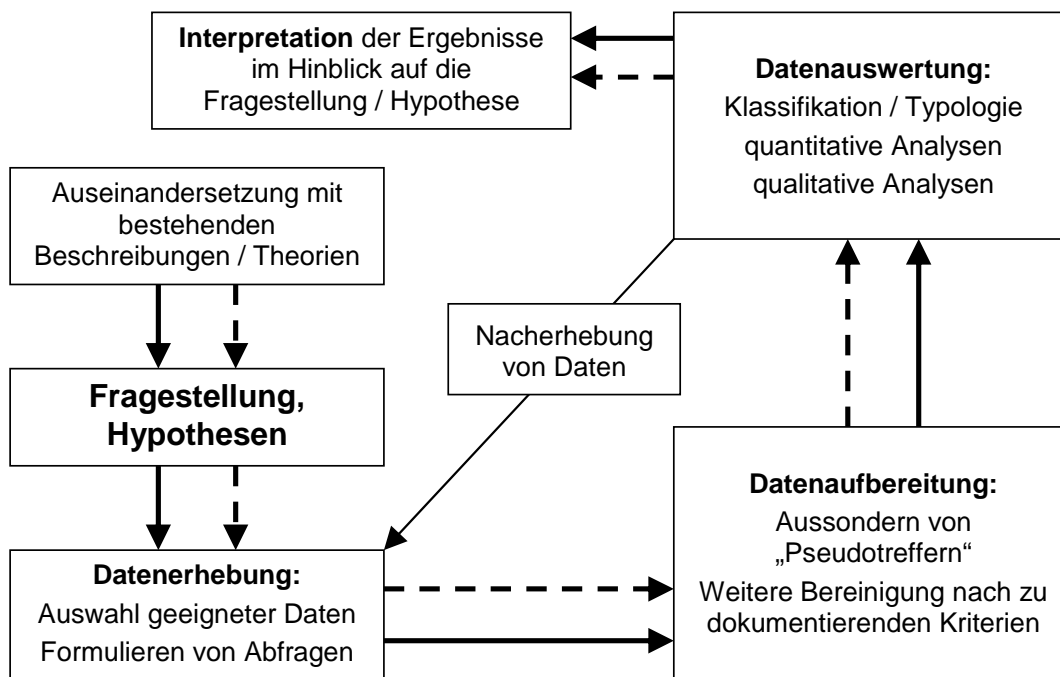


Abbildung 2: Schema für die Planung korpusbasierter Studien.

3.2 Erfahrungen mit Fehlerquellen und Grenzen der aktuellen Korpus-technologie

Im Folgenden möchten wir einige Beispiele für solche Fehlerquellen und Grenzen herausgreifen, die – nicht nur von Studierenden, sondern generell von computerlinguistisch nicht vorgebildeten Korpusnutzern – oft als „negative Überraschung“ empfunden werden und die deshalb aus unserer Erfahrung explizit thematisiert werden sollten:

Grenzen der lemmabasierten Abfrage: Studierende begreifen schnell, warum es in einer flektierenden Sprache wie dem Deutschen von Vorteil ist, in einem Korpusrecherchesystem zwischen einer lemmabasierten und einer formbasierten Suche unterscheiden zu können. Sie gehen allerdings davon aus, dass die Trefferliste einer lemmabasierten Abfrage alle Formen des Paradigmas enthält (optimale Ausbeute) und dass keine nicht zum Paradigma gehörigen Formen als Treffer gelistet werden (optimale Präzision). Leider trifft dies nicht immer zu: Probleme mit der Präzision gibt es beispielsweise bei Lemmata, deren Flexionsparadigmen sich überschneiden (z.B. *raten* und *geraten*; *gelingen* und *gelangen*). Wenn man diesen Falltyp kennt, kann man antizipieren, bei welchen Lemmata Probleme auftreten könnten und Strategien entwickeln, um durch spezifischere Abfragen die Präzision zu erhöhen. Probleme mit der Ausbeute gibt es generell bei den deutschen Partikelverben (*einbringen*, *aufstehen*, *abtreten* etc.): Die lemmabasierte Suche nach Partikelverben in Korpora erfasst nämlich nur die Formen, in denen Partikel und Verb zusammengeschrieben sind (also *einbrachte*, *einggebracht*, *einbrächte*, *einbringst* etc). Belege, in denen Verbpartikel und finiter Verbstamm in getrennter Position die Satzklammer bilden (*brachte ... ein*, *bringst ... ein*) werden in der Trefferliste nicht angezeigt. Auch hier gilt:

Wenn man diese Beschränkung kennt, kann man sie durch geschickte Abfragetechnik kompensieren.

Ob und welche Kompromisse man an Präzision und Ausbeute einer Abfrage stellen möchte, hängt letztlich von der verfolgten Fragestellung ab. Wichtig ist es zunächst, sich überhaupt der Möglichkeit bewusst zu sein, dass in den automatisch erzeugten Trefferlisten Pseudotreffer enthalten sind bzw. relevante Belege nicht gefunden werden. Wer z.B. Vorkommensfrequenzen von Verben vergleicht, muss bei der Interpretation seiner Daten berücksichtigen, dass bei einer lemmabasierten Abfrage zu *geraten* auch die Formen von *raten* mitgezählt werden. Umgekehrt gilt wiederum, dass bei einer lemmabasierten Abfrage zu *bringen* auch eine nicht unerhebliche Anzahl von Treffern für Partikelverben wie *einbringen* oder *abbringen* ausgegeben werden, und zwar genau diejenigen, in denen finiter Verbstamm und Partikel in getrennter Position die Satzklammer bilden. Die automatisch generierten Frequenzen können also erheblich von den um Pseudotreffer bereinigten Frequenzdaten abweichen.

Grenzen der Wortartenannotation: Der Nutzwert der automatischen Wortartenannotation kann gerade am Beispiel der deutschen Partikelverben sehr gut verdeutlicht werden: Wer nach Belegen sucht, in denen Partikel und finiter Verbstamm von *einfinden* in getrennter Position im Satz auftreten, kann die Präzision beträchtlich erhöhen, indem er die Suche auf Verwendungen einschränkt, in denen *ein* als trennbare Verbpartikel annotiert ist. Es ist im Allgemeinen kein Problem, die Studierenden mit den Kategorien der Wortartenannotation, z.B. dem Stuttgart-Tübingen-Tagset STTS, vertraut zu machen.² Viele Studierende sind jedoch enttäuscht, wenn sie feststellen, dass die automatische Wortartenzuordnung nicht fehlerfrei funktioniert und die prinzipiell sinnvoll formulierten Abfragen dennoch weiterhin viele Pseudotreffer (*false positives*) enthalten. Fehleranfällig sind nicht nur die trennbaren Verbpartikeln, sondern auch viele andere homographe Wortformen, die zu unterschiedlichen Wortarten gehören (z.B. *schicke* als Verb- oder Adjektivform, *sein* als Verb- bzw. Pronomenform, *ja* als Responsiv oder als Abtönungspartikel etc.). Um Frustration zu vermeiden, müssen die Probleme der automatischen Verfahren und Ansätze zur künftigen Optimierung transparent gemacht werden.

„Semantische Blindheit“ der Korpusabfragen: Bei vielen linguistischen Analysefragen sucht man eigentlich nach sprachlichen Einheiten in einer bestimmten Bedeutung, also etwa nach *Backfisch* als Bezeichnung für eine Jugendliche oder nach *Ampel* als Bezeichnung für eine Parteienkoalition. Bisher ist eine solche semantische Suche aber in Referenzkorpora nicht möglich, denn sie würde eine automatische Disambiguierung aller mehrdeutigen Wortformen voraussetzen. Davon abgesehen, dass es gar nicht einfach ist, sich auf Kriterien für die Unterscheidung semantischer Lesarten zu einigen – ein Vergleich verschiedener Wörterbücher macht dies deutlich –, gehört die automatische Lesartendisambiguierung (engl. *word sense disambiguation*, *WSD*) leider trotz intensiver Forschungsbemühungen zu den noch nicht befriedigend gelösten Aufgaben in der Computer-

linguistik (vgl. den Überblick in RAYSON & STEVENSON 2008). Man kann deshalb derzeit und ggf. auch noch in absehbarer Zukunft in großen Korpora nicht automatisch nach bestimmten Wortbedeutungen suchen. Über diese „semantische Blindheit“ der aktuellen Korpustechnologie sind gerade Einsteiger in die korpusgestützte Sprachanalyse enttäuscht, denn sie führt bei der Analyse von stark polysemen Wortformen zu einem hohen manuellen Nachbearbeitungsaufwand. Wer häufiger mit Korpora arbeitet, entwickelt jedoch meist bald ein Gespür dafür, welche Fragen mit welchem Aufwand und mit welchen Suchstrategien bearbeitbar sind. Methodisch muss das Problem der semantischen Blindheit gerade bei der Bewertung statistischer Ergebnisse im Auge behalten werden, denn auch die Statistiken operieren über Formeinheiten und nicht etwa über Bedeutungseinheiten. Wer der Frage nachgeht, ob der Ausdruck *Backfisch* vom Ausdruck *Teenager* verdrängt wird, kann sich nicht allein auf automatisch erzeugte Frequenzverlaufdiagramme verlassen, sondern muss zunächst die Belege, in denen „Backfisch“ als Bezeichnung für ein Nahrungsmittel verwendet wird, aussondern.

Insgesamt ist es aus unserer Erfahrung wichtig, den Lehramtstudierenden, die ja computerlinguistisch nicht ausgebildet (und meist auch nicht interessiert) sind, zumindest die grundlegenden Verfahrensweisen der automatischen linguistischen Aufbereitung von Korpusdaten transparent und damit die Problemfelder verstehbar zu machen. Wenn dies gelingt, ist nach unserer Erfahrung auch die Bereitschaft vorhanden, die aktuellen Grenzen zu akzeptieren oder diese sogar als Herausforderung zu sehen, um Heuristiken für optimierte, fehlerrobuste Korpusabfragen zu entwickeln. Bereits an kleinen Praxisbeispielen wird ja schnell deutlich, dass auch die noch nicht perfekte digitale Korpustechnologie Analyseoptionen bietet, die durch eine manuell-intellektuelle Analyse von „traditionellen“ Textkorpora in Papierform nicht oder nur mit extrem hohem Aufwand denkbar wären.

4 Einsatz digitaler Sprachressourcen in Lehrveranstaltungen zum Thema „Internetbasierte Kommunikation“

4.1 Gegenstand und Seminarszenario

Die wissenschaftliche Analyse des Sprachgebrauchs und der kommunikativen Besonderheiten in der internetbasierten Kommunikation ist ein Forschungsschwerpunkt der Dortmunder germanistischen Linguistik. Die Befunde, Beschreibungsansätze und Positionen aus diesem Forschungsbereich sind für Studierende der Lehramter Deutsch sowohl unter fachwissenschaftlicher wie auch fachdidaktischer Perspektive relevant:

(1) Netzgestützte Kommunikationsmöglichkeiten (insbesondere E-Mails, Instant Messaging, Foren, Chats und „soziale Netzwerke“) erfreuen sich in allen Altersgruppen immer stärkerer Beliebtheit, wobei die 14-19-Jährigen die Altersgruppe mit der höchsten Online-Affinität darstellen. Bei der Behandlung sprachlicher Varietäten und sprachlicher Entwicklungstendenzen im Deutschunterricht der Sekundarstufen wird man auf Dauer nicht umhinkommen, diesen Kommunikationsbereich, der gerade für jüngere Nutzergrup-

pen einen selbstverständlichen und wichtigen Teil ihrer alltäglichen Kommunikationswirklichkeit darstellt, systematischer zu berücksichtigen.

(2) Im öffentlichen Diskurs über den Sprachgebrauch im Netz wird häufig die Besorgnis um den Verfall schriftsprachlicher Kompetenzen thematisiert. Wir möchten die Studierenden dazu befähigen, sich in ihrer zukünftigen Rolle als DeutschlehrerInnen und Akteure in staatlichen Bildungsinstitutionen fachkompetent an diesem Diskurs über die Besonderheiten des Sprachgebrauchs im Internet zu beteiligen. Sie sollen in die Lage versetzt werden, die Auffälligkeiten in der internetbasierten Kommunikation in übergreifende Strömungen der Entwicklung von Schriftlichkeit einzuordnen und funktional zu deuten. Sie sollen auch ein Bewusstsein dafür entwickeln, dass stilistische Variation im Sprachgebrauch kein ausschließlich medienabhängiges Phänomen darstellt (etwa im Sinne einer „Netzsprache“ gegenüber der Sprachverwendung außerhalb des Netzes), sondern dass die Sprachverwendung auch *innerhalb* der internetbasierten Kommunikation, abhängig von Zweckbereichen sprachlichen Handelns und unterschiedlichen sozialen/institutionellen Kontexten, z.T. erheblich variiert (z.B. zwischen Chats im Freizeitbereich und Chats in Beratungskontexten, vgl. das Beispiel in Abb. 3).

(3) Darüber hinaus kann und sollte die Thematik auch im schulischen Deutschunterricht (in verschiedenen Klassenstufen) angesprochen werden. Anknüpfungspunkte an die Lehrpläne sind durchaus vorhanden: Der Einfluss technischer Vermittlung auf die Gestaltung von Kommunikationsprozessen ist Gegenstand des Kompetenzziels, „die durch die Informations- und Kommunikationstechnologien bewirkten Veränderungen im Denken, Wahrnehmen und Kommunizieren bewusst“ zu machen, das im Lehrplanbereich „Reflexion über Sprache“ (vgl. RLP Sek II Deutsch NRW: 25) verankert ist. Im Rahmen des generellen Kompetenzziels „Sprachvarietäten untersuchen und angemessen verwenden können“ (RLP Sek II Deutsch NRW: 24) sollten heutzutage auch die unter Punkt (2) angesprochenen schriftsprachlichen Wandelprozesse und sprachlichen Besonderheiten thematisiert werden.

Um der Relevanz der Thematik Rechnung zu tragen, bieten wir an der TU Dortmund regelmäßig fachwissenschaftliche und fachdidaktische Seminare zur internetbasierten Kommunikation an, die im Wahlpflichtbereich der Lehramtsstudiengänge verankert sind. Beispiele für fachwissenschaftliche Seminarthemen sind „Mediale Bedingungen des kommunikativen Handelns“, „Internetbasierte Kommunikation“, „Linguistische Chat-Forschung“ und „Kommunikative Stilistik“; in ihnen werden Konzepte, Forschungsansätze und -ergebnisse zur internetbasierten Kommunikation vermittelt. Im regelmäßig angebotenen fachdidaktischen Seminar „Internetbasierte Kommunikation im Deutschunterricht“ bewerten Studierende vorhandene Unterrichtskonzepte zum Thema bzw. stellen eigene Ideen zur Diskussion.

In beiden Bereichen kommen dabei auch digitale Sprachressourcen zum Einsatz. Im fachdidaktischen Kontext dienen sie vor allem als Materialgrundlage für entsprechende Unterrichtseinheiten. Aus dem weiter unten beschriebenen Dortmunder Chat-Korpus lassen sich beispielsweise Datenbeispiele auch aus solchen kommunikativen Handlungsbe-

reichen im Netz gewinnen, die den Schülerinnen und Schülern aus ihrer alltäglichen Online-Nutzung eher weniger vertraut sein dürften (z.B. Chats in Kontexten institutioneller Beratung oder im Kontext politischer Information und Meinungsbildung). Die Gegenüberstellung von Chats aus unterschiedlichen Handlungsbereichen (vgl. Abb. 3) dokumentiert die große Bandbreite der Sprachlichkeit in Chats und kann als Anknüpfungspunkt dienen, um den Zusammenhang von Sprachstil und Faktoren des kommunikativen Settings herauszuarbeiten und auf dieser Basis die Kompetenz zur situativ angemessenen Sprachwahl auch für die neuen netzbasierten Kommunikationsformen zu stärken (vgl. STORRER 2007).

(a)	
SPOOKY	Irgendwie ist jetzt an mir was vorbeigeschossen
Findalf	Hausdrache, nö, und ja, er ist scheiß langsam!
Arktikus	GFi: *ggg*...hmm..der aiuch...auff jden Fall zu KArneval *s*
desertstorm betritt den Raum.	
<i>ruebennase langweilt sich immer noch....</i>	
GFi	Karneval in Herne? har..
SPOOKY	Hallo ruebennase, wieso langweilst du dich ?
Hausdrache	Hat jemand ne Ahnung, wie ich CarpeDiem per Mail erreiche??
Arktikus	SPOOKY: so froh, daß Du ein Hausgesit bist und kein menschliches Wesen.....sonst wäre das wohl noch insAuge gegangen...:-)
Arktikus	sei froh..solte es heissen
Findalf	spooky, aha und was war das? sah es aus wie text?*g*
ruebennase	spooky, weil keiner mit mir chattet
(b)	
BENUTZER	Können Sie mir sagen, ob das Buch Fn 25665 ausgeliehen ist?
AUSKUNFT	Hallo, wenn Sie einen Moment Geduld haben, schaue ich im Regal nach - Moment
BENUTZER	danke
AUSKUNFT	Ist da, ich lege es Ihnen bei der Information im Erdgeschoss zurück, wenn Sie mirbitte Ihren Namen schreiben.
BENUTZER	Benutzer - bis wann muß ich es abgeholt haben?
AUSKUNFT	Bis wann schaffen Sie es, dann mache ich den entsprechenden Hinweis dran?
BENUTZER	heute oder morgen
AUSKUNFT	O.k. dann schreibe ich bis morgen drauf.
BENUTZER	Vielen Dank!
AUSKUNFT	Gern geschehen und schönen Tag noch.
*** BENUTZER hat den chat verlassen. ***	

Abbildung 3: Gegenüberstellung von Chats aus unterschiedlichen Handlungsbereichen – zum Beispiel (a) „Plauder-Chat“ und (b) chatbasierte Bibliotheksauskunft.

In den fachwissenschaftlichen Seminaren dienen authentische Datenbeispiele und exemplarische korpusgestützte Untersuchungen als Grundlagen für die (selbstständige) linguistische Analyse und Bewertung sprachlicher, interaktionaler und sozialer Besonderheiten bei der Kommunikation mit Online-Medien. Die Studierenden bearbeiten seminarbegleitend kleine Analyseprojekte auf der Grundlage von Sets authentischer Sprachdaten. Die den Analyseprojekten zugrunde liegenden Forschungsfragen und Untersuchungsdesigns werden dabei entweder vorgegeben oder auf der Basis eigener Ideen gemeinsam mit den Studierenden entwickelt und eingegrenzt. Die Projektarbeit wird unterstützt durch eine

Einführung in die genutzten Korpora sowie Hilfestellungen und Beratungsmöglichkeiten durch im Umgang mit den genutzten Ressourcen geschulte TutorInnen. Einzelne Analyseprojekte aus den Seminaren wurden in der Folge im Rahmen von Bachelor-, Master- und Staatsarbeiten ausgebaut, etwa zum Sprachstil und zur sprachlichen Variation in der deutschsprachigen *Wikipedia* (Vergleich von Artikel- und Diskussionsseiten), zur sprachlichen Variation in der Chat-Kommunikation und ihrer Didaktisierung, zu den Funktionen von Nicknames in der internetbasierten Kommunikation, zu Leserkommentaren im Online-Journalismus und zum Phänomen des „Splittings“ im Chat.

4.2 Genutzte Ressourcen / Dortmunder Chat-Korpus

Im Vergleich zu dem in Abschnitt 3 dargestellten Seminarkonzept stehen in den Seminaren zum Bereich „Internetbasierte Kommunikation“ die digitalen Sprachressourcen nicht im Vordergrund, sondern werden punktuell genutzt: als Fundgrube zur Illustration bestimmter Phänomene, als Grundlage für kleine und größere Analyseaufgaben, als empirische Basis für hypothesenerkundende und hypothesenprüfende Untersuchungen zu themenbezogenen Aufgabenstellungen. Da die online verfügbaren Korpusansammlungen zur deutschen Sprache (die Korpusansammlungen am IDS in Mannheim, die DWDS-Korpora) den Bereich der internetbasierten Kommunikation bislang nicht erfassen, müssen dazu Spezialkorpora genutzt (vgl. BEIßWENGER 2007b, BEIßWENGER & STORRER 2008) bzw. eigene Datensammlungen angelegt werden, z.B. zu Leserkommentaren im Online-Journalismus oder zur Beitragsproduktion in Instant-Messaging-Dialogen. Für Analysen zur Interaktionsorganisation in der Chat-Kommunikation wurde in verschiedenen Seminaren weiterhin eine Sammlung von Transkripten zu Nutzeraktivitäten beim Chatten eingesetzt, die auf einem im Rahmen von BEIßWENGER (2007a) erhobenen Set von Screen-Capturing- und Videodaten basiert.³

Da ein Schwerpunkt der Dortmunder Forschung zur internetbasierten Kommunikation im Bereich der Chat-Kommunikation liegt (vgl. BEIßWENGER (Hrsg.) 2001, BEIßWENGER & STORRER (Hrsg.) 2005), wurde in den Jahren 2002–2008 am Lehrstuhl ein Korpus mit Chat-Mitschnitten aufgebaut, das 140.000 Nutzerbeiträge bzw. 1,06 Millionen Tokens aus unterschiedlichen sozialen Handlungsbereichen umfasst: Neben Webchats und IRC-Chats im Freizeitbereich dokumentiert das *Dortmunder Chat-Korpus* Chats in Lehr-/Lernkontexten, in verschiedenen Formen institutioneller Beratung und in journalistischen Nutzungskontexten. Das Korpus wurde nicht nur als empirische Basis für die eigene Forschung (STORRER 2007, LUCKHARDT 2009, BEIßWENGER i.Dr.), sondern auch im Rahmen der o.g. Seminare genutzt.

Die für das Korpus erhobenen Mitschnittsdaten wurden zunächst automatisch bereinigt und in ein einheitliches XHTML-Basisformat konvertiert. Anschließend wurden sie in mehreren semiautomatisch und manuell durchgeführten Aufbereitungsschritten in ein

3 Vier Beispiel-Transkripte aus dem Transkriptkorpus stehen online unter <http://www.michael-beisswenger.de/sprachhandlungskoordination/> zur Verfügung.

XML-Format überführt, das die Struktur der Mitschnitte und der einzelnen Nutzerbeiträge modelliert, die Nutzerbeiträge klassifiziert, systemgenerierte von nutzergenerierten Primärdatensegmenten unterscheidbar macht und ausgewählte Stilelemente internetbasierter Kommunikation auszeichnet. Hierzu gehören z.B. Emotikons und Inflektive sowie Formen, mit denen Chatter ihre Beiträge an andere Chatter adressieren (vgl. zur Adressierung z.B. BEIBWENGER 2000: 79ff.).

Ein 383 Dokumente (~550.000 Tokens) umfassender, teilweise anonymisierter Ausschnitt des Korpus („Releasekorpus“) wird seit 2005 zusammen mit einem spezialisierten, Java-basierten Suchwerkzeug (*STACCADo*) unter <http://www.chatkorpus.tu-dortmund.de> zur freien wissenschaftlichen Nutzung bereitgestellt. Die in das Korpus eingebrachten Annotationen erlauben das Ausblenden von Systemmeldungen bei der Korpusrecherche, die gezielte Suche nach bestimmten (oder innerhalb bestimmter) Typen von Nutzerbeiträgen (Beiträge im Standard-Modus vs. „action messages“) oder auch innerhalb der Beiträge einzelner Nutzer. Darüber hinaus können die im Korpus annotierten Stilelemente automatisch ausgefiltert, gezählt oder mit variabel zugeschnittenen Kotexten ausgegeben werden. Für jeden in einem Mitschnitt als Beitragsproduzent dokumentierten Nutzer sind darüber hinaus verschiedene Typen von Metadaten erfasst, u.a. das vermutete Geschlecht sowie die Anzahl der von ihm produzierten Beiträge und Tokens. Auf Basis dieser Metadaten lassen sich mit *STACCADo* automatisch statistische Übersichten zu einzelnen Teilkorpora oder Korpusdokumenten erzeugen: Sog. „Logfile-Profile“ machen verschiedene Teilkorpora hinsichtlich der durchschnittlichen Länge von Chat-Beiträgen oder der Frequenz der Verwendung ausgewählter Sonderelemente wie z.B. Emotikons vergleichbar; „Chatter-Profile“ liefern Übersichten zu den Anteilen einzelner Chatter am Beitragsaufkommen in den jeweils ausgewählten Teilkorpora sowie zum Verhältnis ihrer durchschnittlichen Beitragslängen zur Beitragslänge eines automatisch errechneten „Durchschnitts-Chatters“ und geben für jeden im ausgewerteten Korpusteil bezeugten Chatter das vermutete Geschlecht aus.

Neben der XML-Version, die zur Nutzung zusammen mit dem Suchwerkzeug *STACCADo* heruntergeladen werden muss, ist der online angebotene Korpusausschnitt auch in einer HTML-Version verfügbar, die die 383 Dokumente direkt im Browser darstellbar macht. Auf diese Weise werden die Daten auch für Lehrkräfte und andere am Thema Interessierte zugänglich, die das Korpus nur als Fundgrube für Unterrichtsmaterial nutzen möchten und keine systematischen Auswertungen anstreben. Das Chat-Korpus wurde in 2009 in das Kerncurriculum Deutsch für die gymnasiale Oberstufe des Landes Niedersachsen als Ressource für den Unterricht im Wahlpflichtmodul 2 „Die deutsche Sprache unter dem Einfluss der Neuen Medien“ aufgenommen (vgl. KC Deutsch GyGe NI: 51). Ein Beispiel für eine Aufgabenbeschreibung zu einem Analyseprojekt auf der Grundlage des Korpus aus dem Seminar „Linguistische Chat-Forschung“ ist in Abb. 4 wiedergegeben.

Forschungsfrage: Variiert der Gebrauch von Interjektionen und Abtönungspartikeln in unterschiedlichen Typen von Chats?

- (1) Orientieren Sie sich in verschiedenen Grammatiken (DUDEN-Grammatik, IDS-Grammatik etc.) über Formen und Funktionen von **Abtönungspartikeln** (wie z.B. *aber, denn, vielleicht, bloß, halt, ...*) und **Interjektionen** (wie z.B. *hm, ah, ach, oh, äh, ...*).
- (2) Suchen Sie (manuell) alle Abtönungspartikeln und Interjektionen, die in den beiden Beispiel-Mitschnitten 1+2 vorkommen, die Sie aus dem Ordner „Materialien“ im Dateibereich der *Stud.IP*-Plattform herunterladen können. **ACHTUNG:** Zu einigen Partikeln und Interjektionen (z.B. *aber, denn, oder, ...*) gibt es homonyme (gleichlautende) Wörter in anderen Wortartenklassen! Erstellen Sie eine Übersicht, welche Interjektionen und Abtönungspartikeln in den beiden Mitschnitten wie oft vorkommen.
- (3) Durchsuchen Sie anschließend (automatisch) mit dem Suchwerkzeug *STACCADO* die folgenden Teilkorpora des Dortmunder Chat-Korpus nach Vorkommnissen derjenigen Wortformen, die Sie in den beiden Beispiel-Mitschnitten als Interjektionen und Abtönungspartikeln vorgefunden haben: (a) alle Plauder-Chats in Medienkontexten; (b) alle Beratungschats der UB Dortmund; (c) die Expertenchats im Hochschulkontext; (d) alle Teilkorpora, die moderierte Chats mit Politikern und Prominenten enthalten.
Berücksichtigen Sie bei der automatischen Suche auch mögliche Schreibvarianten (z.B. *hmhm, oda, ...*).
Lassen Sie sich von *STACCADO* alle Fundstellen in den durchsuchten Teilkorpora inklusive eines Kontextausschnitts von 10 Beiträgen vor und nach der Fundstelle ausgeben, speichern Sie das Ergebnis in einer Datei und eliminieren Sie manuell Pseudotreffer (z.B. Konjunktion *denn*, Adverb *vielleicht, ...*).
- (4) Ermitteln Sie für jede Interjektion / Abtönungspartikel, wie häufig sie in den durchsuchten Teilkorpora insgesamt auftritt (einmal in absoluten Zahlen, einmal in Form einer Angabe zur Frequenz pro 100 nutzer-generierten Tokens).
- (5) Falls die Verteilung in den durchsuchten Teilkorpora unterschiedlich ist, überlegen Sie Gründe, woran das liegen könnte.

Abbildung 4: Analyseaufgabe einer Projektgruppe im Hauptseminar „Linguistische Chat-Forschung“ (SS 2007).

4.3 Erfahrungen und Herausforderungen

Nach unseren Erfahrungen mit daten- bzw. korpusgestützten Analyseprojekten in Lehrveranstaltungen zur internetbasierten Kommunikation stehen Studierende einer selbstständigen Bearbeitung kleiner Forschungsfragen auf der Grundlage authentischer Sprachdaten durchaus aufgeschlossen gegenüber. Die entsprechenden Projekte werden meist durchaus engagiert bearbeitet – vorausgesetzt, der praktische Umgang mit der Nutzung des entsprechenden Korpus oder Datensets wurde zuvor eingeführt und es wurden Hinweise (und Begründungen) zur Bewertung der bei der automatischen Korpusrecherche ermittelten Trefferlisten gegeben.

Als sehr sinnvoll hat sich das Angebot einer Tutoren-Sprechstunde speziell zu technischen Fragen (zum Umgang mit der Korpuschnittstelle, zur Abfragesyntax etc.) erwiesen. Hierfür bedarf es TutorInnen, die einerseits über die erforderlichen technischen Kompetenzen verfügen, andererseits selbst bereits grundlegende Erfahrungen mit der Durchführung korpusgestützter Analyseprojekte gesammelt haben. Mit einer Verstetigung entsprechender Veranstaltungen im Fach können Studierende mit dem erforderlichen Kompetenzprofil aus den Seminaren selbst rekrutiert und in Folgeseminaren als TutorInnen eingesetzt werden. Auch die Möglichkeit, vor der eigentlichen Präsentation der Projektergebnisse im Seminarplenum (z.B. in der Semestermitte) Zwischenergebnisse zu diskutieren und Fragen, die sich bei der praktischen Durchführung der Analysen ergeben haben,

zu besprechen, hat sich bewährt, lässt sich aber i.d.R. nur in kleineren Seminaren sinnvoll realisieren.

Bei den Korpora, die derzeit zum Bereich der internetbasierten Kommunikation existieren, handelt es sich entweder um reine Rohdatenkorpora, deren Inhalte zu Zwecken einer einheitlichen Verwaltung bestenfalls formal vereinheitlicht wurden, ansonsten aber keiner weiteren Aufbereitung speziell in Hinblick auf linguistische Analysezwecke unterzogen wurden (vgl. BEIßWENGER 2007b, BEIßWENGER & STORRER 2008). Die wenigen Korpora, deren Datenbestand entsprechend aufbereitet wurde (z.B. das Dortmunder Chat-Korpus), sind – zumindest zu großen Teilen – *handannotiert*. Im Vergleich zu den in Abschnitt 3 erwähnten linguistisch aufbereiteten Textkorpora gibt es keine Lemmatisierung und auch kein Wortartentagging. Dies hat damit zu tun, dass Werkzeuge für die automatische linguistische Aufbereitung von Sprachdaten (Tokenisierer, Lemmatisierer, Part-of-speech-Tagger, Chunk Parser) meist an redigierter Schriftsprache (oft Zeitungskorpora) trainiert sind und sich bei den wenig normgerechten Chat-Daten nur begrenzt bewähren. Eine Anpassung dieser Werkzeuge an Abweichungen von der (gerade in Zeitungskorpora i.d.R. sehr mustergültig eingehaltenen) orthographischen Norm, die nicht nur für Chats, sondern für viele informelle Nutzungskontexte der internetbasierten Kommunikation typisch sind, muss erst noch geleistet werden – diese Anpassung ist ein wichtiges Desiderat in Bezug auf die Verarbeitung von Sprachdaten aus dem Netz und ihre Repräsentation in Korpora. Entsprechend steht die Möglichkeit einer Disambiguierung formgleicher Ausdrücke durch Einbeziehung etwa von Wortartenkategorien oder eines Grundformenoperators in die Suchanfragen für solche Korpora nicht zur Verfügung. Je nachdem, welche Kategorien im Korpus (semi-automatisch oder manuell) annotiert wurden und welche nicht, muss also sowohl für die Ermittlung der für eine Analysefrage relevanten Treffer wie auch für die intellektuelle Nachsortierung der für die Anfrage faktisch erhaltenen Treffer ein ungleich größerer Aufwand betrieben werden als im Falle der Arbeit mit linguistisch aufbereiteten Textkorpora. Deshalb ist es gerade für die Arbeit mit Spezialkorpora zur internetbasierten Kommunikation wichtig, ein Verständnis dafür zu vermitteln, welche Arten von Fragestellungen sich auf der Grundlage der im Korpus enthaltenen Daten und Annotationen direkt beantworten lassen und für welche Arten von Fragen die ermittelten Treffer nur „rohe“ Ergebnisse darstellen, in deren Bewertung noch weitere intellektuelle Nacharbeit investiert werden muss.

5 Desiderate und Perspektiven

Die Behandlung von und der systematische Umgang mit digitalen Sprachressourcen ist derzeit zwar noch kaum explizit in den Lehrplänen für Schulen und den fächerspezifischen Bestimmungen für die Lehramtsstudiengänge mit Deutsch verankert, lässt sich aber – wie in Abschnitt 2 gezeigt wurde – sehr gut an verschiedene der dort formulierten Kompetenzziele anschließen. Für die Nutzung digitaler lexikalischer Ressourcen zur deutschen

Sprache (Wörterbücher, Wörterbuchportale und digitale lexikalische Informationssysteme) werden allerdings erweiterte Kompetenzen zur Bewertung von Verlässlichkeit und Qualität benötigt, die in der „prädigitalen“ Wörterbuchdidaktik noch kaum thematisiert sind (vgl. BEJOINT 1989). Wie wir in den Abschnitten 3 und 4 an Beispielen gezeigt haben, bieten digitale Textkorpora zwar viele interessante Anknüpfungspunkte für die fachwissenschaftliche und fachdidaktische Ausbildung in den Lehramtsstudiengängen; ob und wie weit die aktuell verfügbaren Online-Korpora zur deutschen Sprache schon direkt im Schulunterricht eingesetzt werden können, muss jedoch noch weiter erprobt werden.

Aus unseren Erfahrungen heraus könnten die folgenden Faktoren dazu beitragen, den Umgang mit digitalen Korpusressourcen auch für computerlinguistisch nicht vorgebildete Nutzer – z.B. im Lehramtsstudium und in der Schule – noch attraktiver zu machen:

Einheitliche Schnittstellen und Anfragesprachen: Bislang muss für fast jede Korpusumgebung eine eigene Abfragesyntax und der Umgang mit einer spezifisch auf dieses Korpus zugeschnittenen Nutzerschnittstelle erlernt werden. Dies erschwert es, Daten aus unterschiedlichen Korpusbeständen zu nutzen oder gar systematisch zu vergleichen (z.B. Text- vs. Gesprächskorpora, Spezialkorpora vs. Referenzkorpora). Ein Desiderat wäre also eine einheitliche Oberfläche für mehrere Korpora sowie die Vereinheitlichung von Anfragesprachen.

Verständliche Dokumentation von Annotationen und von automatisch erzeugten Angaben: Die in ein Korpus eingebrachten Annotationen (Tagsets) und zugrunde liegenden Analysekatoren sollten so dokumentiert sein, dass sie auch für Nutzer ohne computerlinguistische Vorbildung verständlich und nachvollziehbar sind. Wenn in den Projekten die Zeit für eine solche Dokumentation fehlt, sollte man zumindest eine Web2.0-Funktion anbieten, in denen sich die Nutzer wechselseitig austauschen und unterstützen können. Auch die Funktionen, die automatische Angaben aus den Korpusdaten erzeugen (Frequenzangaben, Kookkurrenz- bzw. Wortprofile, Kookkurrenzgraphen), sollten ausführlich dokumentiert werden, um Fehlinterpretationen der Daten zu vermeiden.

Grenzen und Problemfelder offen ansprechen: Grenzen der aktuellen Korpustechnologie sollten in den Hilfetexten und Tutorials zum Umgang mit Korpora offen angesprochen und erklärt werden (z.B. Probleme und typische Fehler bei der automatischen Lemmatisierung und Wortartenannotation oder das Problem der „semantischen Blindheit“ der Korpusuche). Eine Kenntnis der Grenzen automatischer Sprachverarbeitung hilft Nutzerinnen und Nutzern ohne computerlinguistische Grundkenntnisse, die Potenziale von Korpora realistisch einschätzen und für die korpusgestützte Bearbeitung linguistischer Forschungsfragen geeignete Recherchestrategien entwickeln zu können. Wie in Abschnitt 3 erläutert, haben wir die Erfahrung gemacht, dass Studierende diesen Grenzen tolerant gegenüberstehen und sogar Spaß daran haben, nach Abfragen mit möglichst guter Präzision und Ausbeute zu suchen. Voraussetzung ist, dass sie verstehen, wie die Annotationen zustande kommen und warum die automatisch erzeugten Annotationen oft nicht fehlerfrei sind. Wenn Studierende hingegen ohne Einführung eigenständig Korpora nutzen, sind sie

über auftretende Probleme schnell enttäuscht und bewerten die Ressource vorschnell als irrelevant bzw. unbrauchbar.

Hinweise zu Strategien bei der Korpusabfrage: Sehr hilfreich für Nutzerinnen und Nutzer mit keinen oder nur geringen korpuslinguistischen Grundkenntnissen sind Tipps für die Reduktion von Pseudotreffern und für den Umgang mit typischen Problemfällen bei der Korpusrecherche (z.B. orthographische Varianten, Homographen, Partikelverben). Auch diese könnten ggf. über Web2.0-Angebote auch von den Nutzern selbst erstellt werden.

Arbeitsplatz für die Weiterarbeit mit Rechercheergebnissen: Aus den vom Korpusrecherchesystem erzeugten Trefferlisten ergibt sich in aller Regel noch nicht die Antwort auf eine linguistische Forschungsfrage. Um die Trefferliste in eine Liste mit „echten“ Belegen für das gesuchte Phänomen zu überführen, ist es in den allermeisten Fällen erforderlich, die Treffer weiterzubearbeiten, z.B. indem Pseudotreffer eliminiert werden. In weiteren Bearbeitungsschritten können Subklassen gebildet oder weitere Annotationen hinzugefügt werden. Bislang erfolgen diese Schritte der Datenbearbeitung meist unabhängig von der Korpusumgebung (d.h. nach dem Export der Daten in eine Datei) mit einem externen Werkzeug (einer Annotationsumgebung, mit Excel o.ä.). Dadurch lassen sich die weiterbearbeiteten Daten allerdings auch nachträglich i.d.R. nicht mehr ohne Weiteres mit den Korpusdokumenten vernetzen, aus denen sie extrahiert wurden. Häufig ergibt sich aber gerade bei der Weiterarbeit mit einmal aus einem Korpus gewonnenen Treffern die Notwendigkeit, diese erneut mit dem Korpus in Beziehung zu setzen – beispielsweise, um zu einzelnen „problematischen“ Treffern oder Belegen größere Kontexte einzusehen als diejenigen, die beim Export der betreffenden Treffer gewählt oder automatisch mitausgegeben wurden. Ein weiteres Desiderat für die Zukunft wäre deshalb ein „virtueller“ korpuslinguistischer Arbeitsplatz, der die Nutzer nicht nur bei der Abfrage des Korpus und beim Export ihrer Trefferlisten unterstützt, sondern ihnen auch die Weiterarbeit mit den Daten direkt in der Korpusumgebung ermöglicht und/oder Funktionen anbietet, mit denen sich Treffer auch nach ihrer Weiterbearbeitung in externen Werkzeugen wieder im Korpus ermitteln lassen. Ein solcher korpuslinguistischer Arbeitsplatz sollte idealiter auch häufig benötigten Funktionen zur quantitativen Auswertung von Belegensammlungen anbieten. Die Anreicherung von Korpus-Benutzerschnittstellen um solche und weitere Funktionen würde diese schrittweise von Abfrage- und Exportschnittstellen zu *Umgebungen für korpuslinguistisches Arbeiten* weiterentwickeln, in denen das Korpus nicht lediglich als *Ressource* präsentiert, sondern – orientiert an typischen linguistischen Nutzungsszenarien – als *Hilfsmittel* des wissenschaftlichen Prozesses dargeboten wird.

Für die Umsetzung dieser Desiderate wäre es sicherlich hilfreich, wenn die Entwickler von Korpus-technologie, die bislang ja vornehmlich aus dem Bereich der Computerlinguistik oder der linguistisch orientierten Informatik kommen, den Kontakt mit verschiedenen Gruppen von Korpusnutzern intensivieren. Angehende Lehrerinnen und Lehrer sind neben anderen Nutzergruppen (etwa Journalisten, Übersetzern und Sprachmittlern) sicherlich

interessante Multiplikatoren für die vielfältigen Möglichkeiten, die digitale Sprachressourcen und Sprachkorpora bieten.

6 Literatur und erwähnte Sprachressourcen

6.1 Literatur

- ALBERT, RUTH & COR J. KOSTER Koster (2002). *Empirie in Linguistik und Sprachlehrforschung. Ein methodologisches Arbeitsbuch.* Tübingen.
- BEIBWENGER, MICHAEL (2000). *Kommunikation in virtuellen Welten: Sprache, Text und Wirklichkeit.* Stuttgart.
- BEIBWENGER, MICHAEL (2007a). *Sprachhandlungskoordination in der Chat-Kommunikation.* Berlin. New York (Linguistik – Impulse & Tendenzen 26).
- BEIBWENGER, MICHAEL (2007b). *Corpora zur computervermittelten (internetbasierten) Kommunikation.* In: *Zeitschrift für germanistische Linguistik* 35, 496-503.
- BEIBWENGER, MICHAEL (im Druck). *Raumorientierung in der Netzkommunikation. Korpusgestützte Untersuchungen zur lokalen Deixis in Chats.* In: Barbara Frank-Job, Alexander Mehler & Tilmann Sutter (Hrsg.): *Die Dynamik sozialer und sprachlicher Netzwerke.* Wiesbaden.
- BEIBWENGER, MICHAEL (Hrsg., 2001). *Chat-Kommunikation. Sprache, Interaktion, Sozialität & Identität in synchroner computervermittelter Kommunikation. Perspektiven auf ein interdisziplinäres Forschungsfeld.* Stuttgart.
- BEIBWENGER, MICHAEL & ANGELIKA STORRER (2008). *Corpora of Computer-Mediated Communication.* In: Lüdeling & Kytö (eds), 292-308.
- BEIBWENGER, MICHAEL & ANGELIKA STORRER (Hrsg., 2005). *Chat-Kommunikation in Beruf, Bildung und Medien: Konzepte – Werkzeuge – Anwendungsfelder.* Stuttgart.
- BEJOINT, HENRY (1989). *The Teaching of the Dictionary Use: Present State and Future Tasks.* In: Franz Josef Hausmann, Oskar Reichmann, et al. (Hgg.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie, 1. Teilband.* Berlin. New York (HSK 5.1), 208-215.
- BORTZ, JÜRGEN & NICOLA DÖRING (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. 4., überarb. Aufl.* Berlin u.a
- [FSB GERM BFP DORTMUND] *Fächerspezifische Bestimmung für das Fach Germanistik zur Prüfungsordnung für den Bachelor-Studiengang mit fachwissenschaftlichem Profil im Modellversuch „Gestufte Studiengänge in der Lehrerbildung“.* Amtliche Mitteilungen der Universität Dortmund 12/2007, 34-43. URL: http://www.tu-dortmund.de/uni/studierende/pruefungsangelegenheiten/ord/FSB/FSB_Bachelor/15_FSB_Germanistik_BfP.pdf.
- ENGELBERG, STEFAN & LOTHAR LEMNITZER (2009). *Lexikographie und Wörterbuchbenutzung. 4. überarbeitete Auflage.* Stauffenburg: Tübingen.
- EISENBERG, PETER & WOLFGANG MENZEL (1999). *Grammatik-Werkstatt.* In: *Praxis Deutsch* 22, Heft 129, 14-26.
- GEYKEN, ALEXANDER (2007). *The DWDS corpus: A reference corpus for the German language of the 20th century.* In: Christiane Fellbaum (ed.): *Collocations and Idioms,* London, 23-40.

- HOFFMANN, LUDGER (2006). Funktionaler Grammatikunterricht. In: Tabea Becker & Corinna Peschel (Hrsg.): Gesteuerter und ungesteuerter Grammatikerwerb. Baltmannsweiler, 20-45.
- [KC DEUTSCH GYGE NI] Kerncurriculum Deutsch für das Gymnasium/gymnasiale Oberstufe, die Gesamtschule/gymnasiale Oberstufe, das Fachgymnasium, das Abendgymnasium, das Kolleg. Hrsg. v. Niedersächsischen Kultusministerium 2009. WWW-Ressource: http://db2.nibis.de/1db/cuvo/datei/kc_deutsch_go_i_2009.pdf.
- LEMNITZER, LOTHAR & HEIKE ZINSMEISTER (2006). Korpuslinguistik: Eine Einführung. Tübingen.
- LUCKHARDT, KRISTIN (2009). Stilanalysen zur Chat-Kommunikation. Eine korpusgestützte Untersuchung am Beispiel eines medialen Chats. Diss., TU Dortmund. Digitale Ressource: <http://hdl.handle.net/2003/26055>.
- LÜDELING, ANKE & MERJA Kytö (Eds., 2008/2009). Corpus Linguistics. An International Handbook. 2 Bde. Berlin. New York (HSK 29.1/29.2).
- MCENERY, TONY, RICHARD XIAO & YUKIO Tono (2006). Corpus-Based Language Studies – an advanced resource book. London. New York.
- RAYSON, PAUL & MARK STEVENSON (2008). Sense and semantic tagging. In: Lüdeling & Kytö (eds.), 564-578.
- [RLP SEK II DEUTSCH NRW] Richtlinien und Lehrpläne für die Sekundarstufe II – Gymnasium/Gesamtschule in Nordrhein-Westfalen. Deutsch. Hrsg. v. Ministerium für Schule und Weiterbildung, Wissenschaft und Forschung 1999.
- STORRER, ANGELIKA (2007). Chat-Kommunikation in Beruf und Weiterbildung. In: Der Deutschunterricht 1/2007, 49-61.
- STORRER, ANGELIKA (2010). Deutsche Internet-Wörterbücher: Ein Überblick. In: Lexicographica, 155–164.
- STORRER, ANGELIKA (2011). Korpusgestützte Sprachanalyse in Lexikographie und Phraseologie. In: Karlfried Knapp u.a. (Hrsg.): Angewandte Linguistik. Ein Lehrbuch. 3. Auflage. Tübingen, 216-239. [Aufgaben und Lösungen zum Artikel online unter http://www.studiger.tu-dortmund.de/images/Korpuslinguistik_aufgaben.pdf].

6.2 Erwähnte Ressourcen

- [CANOO.NET]: Portal zur deutschen Sprache (Wörterbücher und Grammatik): <http://www.canoo.net/>
- [DICT.CC] Verbund multilingualer Wörterbücher mit Deutsch als Äquivalentsprache: <http://browse.dict.cc/>
- DORTMUNDER CHAT-KORPUS: <http://www.chatkorpus.tu-dortmund.de>
- [DWDS]: Verbund digitaler Wörterbücher und Textkorpora (Berlin-Brandenburgische Akademie der Wissenschaften): <http://www.dwds.de>
- [DWDS-WDG]: Digitales Wörterbuch der deutschen Sprache auf der Basis des digitalisierten „Wörterbuch der deutschen Gegenwartssprache“ (Berlin-Brandenburgische Akademie der Wissenschaften): <http://www.dwds.de>
- [DWB-ONLINE]: Das Deutsche Wörterbuch von Jacob und Wilhelm Grimm Online: <http://www.dwb.uni-trier.de>
- [ELEXIKO]: elexiko – Online-Wörterbuch zur deutschen Gegenwartssprache: <http://www.ids-mannheim.de/lexik/elexiko/>

[GRAMMIS]: Grammis – das grammatische Informationssystem des Instituts für deutsche Sprache: <http://hypermedia.ids-mannheim.de/grammis/>

[OPENTHESAURUS]: Wörterbuch für Synonymie und Assoziationen (deutsch): <http://www.openthesaurus.de/>

[OWID]: Verbund digitaler Wörterbücher (Institut für deutsche Sprache): <http://www.owid.de/>

[PROGR@MM]: ProGr@mm – die propädeutische Grammatik: <http://hypermedia.ids-mannheim.de/programm/>

[SZENESPRACHENWIKI]: Duden: Neues Wörterbuch der Szenesprachen: <http://szenesprachenwiki.de/>

[WIKTIONARY]: Wiktionary: das freie Wörterbuch (deutsch): <http://de.wiktionary.org/wiki/Wiktionary:Hauptseite>

[WORTWARTE]: Die Wortwarte: <http://www.wortwarte.de/>