

Sprachstil und Sprachvariation in sozialen Netzwerken

Angelika Storrer

Preprint. Erscheint in: Barbara Frank-Job, Alexander Mehler & Tilmann Sutter (Hrsg.): Die Dynamik sozialer und sprachlicher Netzwerke. Konzepte, Methoden und empirische Untersuchungen an Beispielen des WWW. Wiesbaden: VS Verlag für Sozialwissenschaften.

1. Einleitung
2. Linguistischer Hintergrund: Schriftlichkeit und sprachlicher Ausbau
 - 2.1 Internetbasierte Kommunikation im Spannungsfeld von Mündlichkeit und Schriftlichkeit
 - 2.2 Interaktionsorientiertes vs. textorientiertes Schreiben
 - 2.3 Sprachvariation und Stilmarker in der internetbasierten Kommunikation
3. Sprachstil und Sprachvariation im Dortmunder Chat-Korpus
 - 3.1 Motivation und Fragestellungen
 - 3.2 Datenbasis: *Das Dortmunder Chat-Korpus*
 - 3.3 Sprachvariation in Abhängigkeit von Handlungsbereichen und Kommunikationsphasen
 - 3.4 Individuelle Sprachvariation und stilistische Präferenzen
4. Sprachvariation in der Wikipedia
 - 4.1 Motivation und Fragestellung
 - 4.2 Artikelseiten vs. Diskussionsseiten in der Wikipedia
 - 4.3 Datengrundlage und Auswahl der analysierten Stilmarker
 - 4.4 Datenauswertung und Interpretation der Ergebnisse
5. Stilmarker der internetbasierten Kommunikation in der journalistischen Prosa
6. Fazit und Ausblick
7. Literatur

1. Einleitung

Schriftsprachliche Kommunikation spielt in sozialen Netzwerken eine zentrale Rolle. Die technisierten Kommunikationsplattformen des Web 2.0 bieten viele neue Schreibanlässe, in denen Schriftsprache nicht nur für den Informationsaustausch oder die kollaborative Zusammenarbeit genutzt wird, sondern auch wichtige Funktionen für den Beziehungsaufbau und die Beziehungspflege übernimmt. Damit wird Schrift erstmals im großen Stil auch in Bereichen der Alltagskommunikation genutzt, im dialogischen, spontanen und informellen Austausch, der bislang eher eine Domäne der gesprochenen Sprache war. Hierbei bilden sich neue Haltungen zu Schreiben und Schriftlichkeit heraus, für die sich charakteristische Sprachgestaltungs- und Formulierungsmuster entwickeln.

In der sprachbesorgten Öffentlichkeit und in den Medien werden diese Entwicklungen oft skeptisch betrachtet (vgl. Brommer 2007; Dürscheid und Brommer 2009). Schließlich weichen Produkte der schriftlichen Netzkommunikation, wie die in Abbildung 1, 3 und 7 gezeigt

ten, deutlich von den normativen Erwartungen ab, die man normalerweise an elaborierte Schriftlichkeit anlegt. In der linguistischen Forschung hingegen werden die Veränderungen recht einheitlich als Ausdifferenzierung der Möglichkeiten schriftsprachlichen Handelns gedeutet, die an die neuen Anforderungen der Kommunikation in sozialen Netzwerken angepasst werden. Diese linguistische Sichtweise soll in Abschnitt 2 ausführlicher begründet werden. Dass diese Sichtweise in der sprachbesorgten Öffentlichkeit oft auf ungläubiges Staunen stößt, liegt auch daran, dass Prozesse des Sprachwandels und die Vielfalt sprachlicher Variationen nur wenigen Sprachteilnehmern bewusst sind. Sicherlich werden sich die Vorbehalte gegen die neuen Schreibformen auch nicht durch empirische Befunde alleine ausräumen lassen. Dennoch kann die empirisch arbeitende Sprachwissenschaft zur Versachlichung der Diskussionen beitragen, indem sie die gegenwärtigen Entwicklungen datengestützt untersucht und in größere sprach- und mediengeschichtliche Zusammenhänge einordnet. Diese Versachlichung ist angesichts der wachsenden Bedeutung von sozialen Netzwerken nicht nur im Hinblick auf die sprachkritische Diskussion in der Öffentlichkeit, sondern vor allem auch für die Thematisierung von Sprachstil und Sprachvariation im schulischen Deutschunterricht ein Desiderat: Schließlich gilt es auch im sprachdidaktischen Kontext, die neuen Schreib- und Stilformen dezidiert von anderen Formen elaborierter Schriftlichkeit abzugrenzen und damit die Kompetenz zur situativ und kontextuell angemessenen Sprach- und Stilwahl zu fördern.

In diesem Kontext der Versachlichung stehen auch die in diesem Beitrag beschriebenen korpusgestützten Untersuchungen. Der korpuslinguistische Ansatz ergänzt empirische Studien zur Entwicklung schulischer Schreibkompetenzen (Sieber 1998; Dürscheid et al. 2010), linguistische Modelle für internetbasierte Schreibprozesse (Beißwenger 2007) und explorative Studien, die Einzelaspekte und spezielle Sprachphänomene an qualitativen Analysen mit kleineren Datensets herausarbeiten (z.B. Günthner und Schmidt 2002; Hoffmann 2004; Androutsopoulos und Ziegler 2003; Schlobinski 2006). Der Schwerpunkt der in diesem Artikel vorgestellten Untersuchungen liegt auf der quantitativen Analyse größerer, texttechnologisch aufbereiteter Datenbestände mit Hilfe von korpuslinguistischen Werkzeugen, die eine automatische Suche und Auswertung der Daten, Annotationen und Metadaten unterstützen (vgl. auch Beißwenger in diesem Band, Diewald und Frank-Job in diesem Band). Am Beispiel der Untersuchungen soll deutlich werden, welchen Beitrag korpuslinguistische Studien zur empirischen Erforschung des Feldes leisten können, welche Erkenntnisse man mit welchem Aufwand gewinnen kann und welche Desiderate sich für den Aufbau bzw. die Weiterentwicklung entsprechender Korpora und Werkzeuge gewinnen lassen.

2. Linguistischer Hintergrund: Schriftlichkeit und sprachlicher Ausbau

In diesem Abschnitt sollen die Kernpunkte der linguistischen Positionen zur Sprachlichkeit in der internetbasierten Kommunikation skizziert werden, vor deren Hintergrund die korpusbasierten Untersuchungen durchgeführt wurden. Im Folgenden werde ich zunächst skizzieren, wie sich die neuen Schreibformen in sprachgeschichtliche Entwicklungen von mündlichen und schriftlichen Diskurstraditionen einordnen lassen. Auf dieser Basis kann man erklären, warum viele Sprachwissenschaftler die sprachlichen Besonderheiten der internetbasierten Kommunikation nicht als Indiz für den Verfall von Sprach- und Schriftkultur bewerten, sondern als Ausbau der Schriftsprache im Hinblick auf neue Anforderungen. Diese Anforderungen stehen im Zusammenhang mit der Nutzung von Schrift für die direkte, dialogische Interaktion in einer technisierten Schreibumgebung, für die ich die Bezeichnung „interaktionsorientiertes Schreiben“ einführen möchte. Im Gegensatz zum *textorientierten* Schreiben, das die

schulische Schreiberziehung bislang prägt, steht beim interaktionsorientierten Schreiben nicht das Schreibprodukt, sondern die laufende Interaktion zwischen den Akteuren des Netzwerks im Zentrum. Für den kommunikativen Erfolg beim interaktionsorientierten Schreiben sind deshalb andere Kompetenzen erforderlich als für das Verfassen von Texten. Aus diesem Grund ist es auch nicht zielführend, die Qualitätsmaßstäbe, die für elaborierte Schrifttexte entwickelt wurden, einfach auf die beim interaktionsorientierten Schreiben entstehenden Produkte zu übertragen. Gerade aus schreibdidaktischer Perspektive ist es vielmehr entscheidend, Kompetenzen für beide Formen des schriftsprachlichen Handelns zu entwickeln (vgl. Storrer i.Dr.).

2.1 Internetbasierte Kommunikation im Spannungsfeld von Mündlichkeit und Schriftlichkeit

Zentral für die linguistische Diskussion ist das Verhältnis der Kategorien „Text“ und „Gespräch“. Da Texte und Gespräche mit teilweise unterschiedlichen Kategorien und Methoden¹ untersucht werden, wird der Status neuen Sprachformen als Texte oder als Gespräche in der Linguistik intensiv diskutiert². Dabei spielt auch das Verhältnis von Mündlichkeit und Schriftlichkeit eine wichtige Rolle. Schon frühe Untersuchungen (Feldweg et al. 1995; Haase et al. 1997) stellten heraus, dass in der internetbasierten Kommunikation viele Merkmale der gesprochenen Alltagssprache mit schriftsprachlichen Mitteln nachgebildet und dass mündliche und schriftsprachliche Versprachlichungsmuster auf neuartige Weise kombiniert werden. Zur Beschreibung und Einordnung dieser Beobachtungen hat man in der deutschen Forschung schon früh³ an terminologische Differenzierungen angeknüpft, die in Arbeiten von Peter Koch und Wulf Oesterreicher⁴ eingeführt wurden und auf Arbeiten des Freiburger SFB „Übergänge und Spannungsfelder zwischen Mündlichkeit und Schriftlichkeit“ zurückgehen. Kern dieses Ansatzes ist die terminologische Separierung von zwei Aspekten, die in der Diskussion um die Opposition „Mündlichkeit“ vs. „Schriftlichkeit“ eine Rolle spielen.

- Der eine Aspekt betrifft die Realisierung sprachlicher Äußerungen als ‚phonisch‘ oder ‚graphisch‘ – im Ansatz als „Medium“ bezeichnet. Für diese Opposition wurden die Termini *medial mündlich* vs. *medial schriftlich* eingeführt.
- Der andere Aspekt bezieht sich auf den sprachlichen Duktus, den man intuitiv mit mündlichen Gesprächen auf der einen und elaborierten Schrifttexten auf der anderen Seite in Verbindung bringt – im Ansatz als „Konzeption“ bezeichnet. Für diese Opposition wurden die Termini *konzeptionell mündlich* vs. *konzeptionell schriftlich* eingeführt.

Die Separierung der beiden Aspekte ist nicht neu⁵; der neue Gedanke, den Koch und Oesterreicher ins Spiel brachten, betrifft den konzeptionellen Aspekt: Im Gegensatz zu dem als dichotomisch eingestuften Oppositionspaar „phonisch – graphisch“ wird die konzeptionelle

¹ Einen umfassenden Einblick in die Unterschiede geben die beiden einschlägigen Handbücher zur Textlinguistik Brinker et al. 2000 und zur Gesprächslinguistik Brinker et al. 2001.

² Vgl. zu diesem Thema z.B. Storrer 2001; Hoffmann 2004; Beißwenger 2007; Herring 2010.

³ Z.B. Haase et al. 1997; Runkehl et al. 1998; Beißwenger 2000; Storrer 2001; im Hinblick auf entsprechende Entwicklungen im Französischen vgl. Thaler 2003, Frank-Job 2009/2010.

⁴ Vgl. Koch und Oesterreicher 1994, 2007.

⁵ Sie gehen zurück auf die von Ludwig Söll eingeführte Unterscheidung zwischen *code parlé* und *code écrit* (vgl. Koch und Oesterreicher 1994: 587). Raible (1999: 11ff.) zieht weitere Verbindungslinien zu Karl Bühlers Unterscheidung von Sprechhandlung (*energeia*) vs. Sprachwerk (*ergon*).

Opposition als Kontinuum modelliert, das verschiedene Abstufungen zwischen den Polen Nähe und Distanz umfasst. Das Kontinuum zwischen Nähe und Distanz steht nach Auffassung von Koch und Oesterreicher (1994: 588) für „anthropologisch begründbare, universale Kommunikationshaltungen“ mit speziellen Kommunikationsbedingungen und dafür typischen Versprachlichungsstrategien. In der kritischen Rezeption des Ansatzes wurde das Verhältnis zwischen Kommunikationsbedingungen und Versprachlichungsstrategien weiter präzisiert.⁶ Die grundlegende Idee der Assoziierung der Nähe (in verschiedenen Dimensionen: räumlich, zeitlich, referentiell, emotional) mit konzeptionell mündlichen Sprachmerkmalen einerseits und der Distanz mit konzeptioneller Schriftlichkeit andererseits blieb als Grundidee aber erhalten. Die Separierung von Nähe und Distanz hat zum Ziel, Prozesse des sprachlichen Wandels und des Ausbaus von Varietätenräumen zu systematisieren. Die begriffliche Differenzierung steht im größeren Kontext eines Ansatzes, der Sprachgeschichte als Geschichte von Varietäten- und Kommunikationsräumen beschreibt, wobei die Autoren davon ausgehen, dass Nähe und Distanz das zentrale Organisationsprinzip bilden, nach dem sich einzelsprachliche Varietätenräume – falls mehrere Sprachen im Spiel sind auch ganze Kommunikationsräume – organisieren.

Auf dieser Basis dieses Ansatzes können die Besonderheiten der neuen internetbasierten Kommunikationspraxen in sprachhistorische Entwicklungen bei der Ausdifferenzierung medial mündlicher und schriftlicher Diskurstraditionen⁷ eingeordnet werden. Wichtig hierfür ist die Erkenntnis, dass die Ausbildung einer konzeptionell elaborierten Schriftsprache mit Prozessen einhergeht, für die Koch und Oesterreicher die von Heinz Kloss eingeführte Metapher des Ausbaus übernehmen. Koch und Oesterreicher (1994: 589ff.) unterschieden dabei zwischen Prozessen des extensiven und des intensiven Ausbaus. Mit *extensivem Ausbau* ist gemeint, dass eine Sprache in immer vielfältigeren schriftbasierten Diskurstraditionen des Distanzbereichs verwendet wird, also etwa in den Bereichen Rechtssprechung, Religion und Wissenschaft. Damit einhergehend entwickelt sich das Inventar der Ausdrucksmittel einer Einzelsprache im Hinblick auf die Anforderungen der schriftlichen Distanzkommunikation weiter. Dieser Prozess wird als *intensiver Ausbau* bezeichnet und betrifft verschiedene Ebenen und Phänomenbereiche, z.B. die Erweiterung des Wortschatzes im Bereich der Abstrakta, aber auch bei den Konnektoren und Präpositionen, die Ausdifferenzierung des Tempus- und Modusgebrauchs sowie die Ausbildung des sog. Nominalstils.

Das Deutsche der Gegenwart ist im konzeptionell schriftlichen Bereich sehr gut ausdifferenziert. Die Bewertung der neuen Sprachformen in der internetbasierten Kommunikation kann nicht isoliert, sondern nur vor dem Hintergrund genereller Veränderungen im Stellenwert der Schriftlichkeit erfolgen. Im Laufe des 20. Jahrhunderts wurde eine Epoche starker Schriftlichkeit durch eine teilweise Reoralisierung der öffentlichen Kommunikation abgelöst (v. Polenz 1999: 37ff.). Elektronische Medien – insbesondere Rundfunk und Fernsehen – haben den extensiven Ausbau durch neue Diskurstraditionen sowohl im medial schriftlichen als auch im medial mündlichen Bereich beschleunigt, wobei die Normen für die schriftliche Distanzkommunikation auch das medialisierte Sprechen in der Distanz beeinflussen. Das Deutsche verfügt über kodifizierte Normen, die sowohl im medial mündlichen als auch im medial schriftlichen Bereich die Sprachlichkeit in der Distanzkommunikation prägen – natürlich immer im Rahmen der jeweils individuellen Sprachfähigkeiten. Computer und Internet haben die Mög-

⁶ Z.B. Koch und Oesterreicher (2007), Ägel und Hennig (2006, 2010).

⁷ Der Ausdruck „Diskurstradition“ steht in diesem Ansatz für Traditionen des Strukturierens und Formulierens, die sich teilweise übereinzelsprachlich entwickeln und wissenschaftlich mit Begriffen wie „Textsorten“, „Gattungen“, „Genres“, „Stile“ erfasst werden (vgl. Koch und Oesterreicher 1994: 593ff.).

lichkeiten, in räumlicher Distanz zu kommunizieren, nun erheblich erweitert. Unter der Perspektive der Sprach- und Varietätengeschichte kann man diese neuesten Entwicklungen als Fortsetzung des extensiven Ausbaus einordnen, als Weiterentwicklung von bestehenden und Herausbildung von neuen Kommunikationsformen im Internet, die in der einschlägigen Forschung als „digitale Genres“ (Mehler et al. 2010) beschrieben werden.

Welchen Einfluss die verstärkte Nutzung internetbasierter Kommunikationsformen auf den Sprachgebrauch in anderen Bereichen hat, lässt sich empirisch nur bedingt ermitteln: Medialer und nicht-medialer Sprachgebrauch sind eng miteinander verwoben, sodass der Sprachgebrauch in einem bestimmten Medium nur in Laborsituationen isoliert untersucht werden kann (vgl. Schmitz 2004: 124). In jedem Fall bringt das Internet eine neue Qualität in die Entwicklung insofern, als es von seiner Übertragungsschnelligkeit her Dialogizität in medialer Schriftlichkeit in qualitativ neuer Weise fördert. Damit wird mediale Schriftlichkeit zunehmend in Handlungsbereichen genutzt, die bisher eher der gesprochenen – mündlichen bzw. fernmündlichen – Sprache vorbehalten waren: im spontanen dialogischen Austausch teilweise zwischen Unbekannten, in sozialen Netzwerken aber auch zwischen Menschen, die sich persönlich kennen und einander nahe stehen. Es entwickeln sich neue Kommunikationsformen mit Parameterprofilen und Sprachmerkmalen in einem bislang noch nicht so stark abgedeckten Bereich, nämlich am Nähepol im Bereich der medialen Schriftlichkeit. Hierfür bilden sich nun neue Muster im Umgang mit Schriftsprache heraus, die auf die Konstellation der Nähekommunikation – situative Einbettung, geringe Planung, Dialogizität, Emotionalität – hin optimiert werden. Der intensive Ausbau erstreckt sich v.a. auf den Bereich der interaktiven Einheiten⁸, die zur Emotionalisierung, zur Kommentierung sprachlicher Handlungen und zur Gesprächssteuerung dienen. Hierzu rechnet man auch die Emotikons und Aktionswörter, die in den folgenden Untersuchungen als Stilmarker für den sprachlichen Duktus in der internetbasierten Kommunikation gewertet werden (s.u.). Wichtig für die sprachkritische Diskussion ist außerdem, dass nicht die Kommunikationstechnologie alleine, sondern ein Zusammenspiel verschiedener sozialer, individuell-situativer oder technischer Faktoren den Sprachstil bestimmt. Anders gesagt: Eine Kommunikationstechnologie wie der Chat ist zwar prädestiniert für konzeptionell mündlich geprägte Szenarien und die entsprechenden Sprachmerkmale. In Kontexten der beruflichen und massenmedialen Chatnutzung kann sich der Sprachstil aber wieder an schriftsprachlichen Normen orientieren (vgl. Abschnitt 3.3).

2.2 Interaktionsorientiertes vs. textorientiertes Schreiben

Um die Unterscheidung von Nähe und Distanz mit den verschiedenen Schreibformen in Verbindung zu bringen, differenziere ich zwischen dem *interaktionsorientierten* Schreiben, das sich in seinen Kommunikationsbedingungen und Versprachlichungsstrategien eher am Nähepol orientiert, und dem *textorientierten* Schreiben, das in seinen vielfältigen Textsorten ein breites Spektrum im Spannungsfeld von (konzeptioneller) Schriftlichkeit und Mündlichkeit abdeckt:

Beim **interaktionsorientierten Schreiben** steht das aktuelle Kommunikationsgeschehen im Fokus der Aufmerksamkeit eines Netzwerk-Akteurs, der in der Interaktion mit anderen stets zwischen Produzenten- und Rezipientenrolle wechselt und auf die Beiträge anderer Akteure

⁸ Zur grammatischen Kategorie „interaktive Einheit“ vgl. die „Grammatik der deutschen Sprache“ (Zifonun et al. 1997, insb. Kap C4.2).

reagiert⁹. Bestimmend für die Versprachlichungsstrategien ist der kommunikative Erfolg in der laufenden Interaktion; hierfür kann die Schnelligkeit der Reaktion wichtiger sein als Prägnanz und sprachliche Elaboriertheit. Für das Interaktionsmanagement und die Beziehungsgestaltung bilden sich neue Formulierungstraditionen und grafische Konventionen heraus, die paraverbale und körpergebundene Kommunikationssignale aufgreifen und in neuer Weise medial realisieren¹⁰. Charakteristisch für viele Produkte des interaktionsorientierten Schreibens ist es, dass sie nur für das Funktionieren in der aktuellen Situation konzipiert sind und nicht für das nachträgliche erneute und mehrfache Lesen. Gerade wegen ihrer neuartigen interaktiven Einheiten eignen sie sich auch nicht für das nachträgliche Vorlesen; sie bleiben – trotz ihres mündlichen Duktus – sehr stark an ihre graphische Form gebunden¹¹.

Beim **textorientierten Schreiben** ist das Schreibziel ein schriftsprachliches Produkt, das über den laufenden Kommunikationszusammenhang hinausgehend Bestand haben soll. Charakteristisch für das textorientierte Schreiben ist die „Dissoziation der Sprechsituation durch den schriftlichen Text“ (Ehlich 1994: 20): Textorientierte Schreibprodukte sind auf eine Rezeptionssituation hin konzipiert, bei denen keine unmittelbare Interaktion zwischen Schreiber und Leser vorgesehen ist; das geschriebene Wort muss aus sich heraus und unabhängig vom kommunikativen Kontext interpretierbar sein. Von den schriftsprachlichen Produkten beim textorientierten Schreiben wird normalerweise erwartet, dass sie im Hinblick auf orthographische Normen und grammatische Standards redigiert sind¹²; außerdem erwartet man eine Anpassung an Strukturierungsmuster und Formulierungstraditionen, die in der Linguistik als „Textsorten“, „Diskursformen“ und „Genres“ systematisch gefasst und beschrieben sind. Das textorientierte Schreiben steht im Zentrum der schulischen Schreibdidaktik (vgl. Storrer i.Dr.); es prägt aber auch viele Schreibenanlässe in beruflichen und institutionellen Kontexten – auch das literarische, journalistische und wissenschaftliche Schreiben ist üblicherweise textorientiert. Die für das textorientierte Schreiben charakteristischen Prozesse des Planens und Überarbeitens, die „Arbeit am Text“, sind in der Schreibprozessforschung inzwischen gut erforscht und beschrieben. In diesem Zusammenhang wird auch der „epistemische Effekt des Schreibens“ hervorgehoben, der darin besteht, dass die mit dem Schreiben verbundene Aufgabe, Ideen und Gedanken in eine visuell sichtbare Ordnung zu bringen, auch beim Schreiben zu einer Klärung gedanklicher Zusammenhänge führt (vgl. z.B. Raible 1999: 42).

2.3 Sprachvariation und Stilmarker in der internetbasierten Kommunikation

Dass die neuen Kommunikationsformen und Versprachlichungsmuster als sprachlicher Ausbau gewertet werden, impliziert, dass sie den allgemeinen Trend zur Erweiterung und Ausdifferenzierung von Diskursformen fortsetzen, während andere etablierte Diskursformen und Versprachlichungsmuster fortbestehen und sich parallel weiterentwickeln. In den sozialen Netzwerken bilden sich dabei sehr viele unterschiedliche Diskursformen heraus, die sich durch technische Kommunikationsfunktionen, soziale Konstellationen und Lebensweltbezüge erheblich unterscheiden. Was in der Linguistik unter Etiketten wie „Netzsprache“, „Sprache

⁹ Vgl. hierzu auch die Beobachtungsstudien in Beißwenger (2007: Kap. 5 u. 6) und (2010).

¹⁰ Vgl. z.B. Runkehl et al. (1998); Beißwenger (2000); Kilian (2001); Günthner und Schmidt (2002); Luckhardt (2009).

¹¹ Vgl. Storrer (2001: 452ff.).

¹² Natürlich lässt das literarische und auch journalistische Schreiben viel Spielraum für Abweichungen von Normen und Standards, diese Abweichungen werden aber als bewusst gesetzte Mittel zur Erzeugung stilistischer Sinns gedeutet.

des Internets“ oder „Internetslang“ beschrieben wird, ist deshalb keine einheitliche Varietät. Vielmehr gibt es nicht nur erhebliche Unterschiede zwischen Kommunikationsformen wie z.B. Chatten und Bloggen, sondern auch eine große Variation innerhalb derselben Form in unterschiedlichen Kontexten. Dies wird in Abschnitt 3 am Beispiel der Sprach- und Stilvariation des Chats gezeigt werden. Dennoch gibt es bestimmte Formulierungsmuster, die als „typische“ Elemente der internetbasierten Kommunikation (bzw. teilweise auch der „Sprache der SMS“) gelten. Dazu gehören Emotikons wie „:-)“ und interaktive Einheiten, die als Akronyme (z.B. „lol“ aus „laughing out loud“) oder aus unflektierten Verbstämmen (sog. Inflektive wie „lach“, „seufz“) gebildet sind. Zwar erschöpfen sich die sprachlichen Besonderheiten der internetbasierten Kommunikation keinesfalls in der Verwendung dieser Elemente; sie sind aber in fast allen einschlägigen Aufstellungen zur „Netzsprache“ bzw. zum „Netzjargon“ aufgeführt und spielen auch für die öffentliche Debatte zum Thema eine zentrale Rolle. Exemplarisch sei ein Ausschnitt aus einem Diskussionsbeitrag zitiert, der unter der Überschrift „Grins, lol, rofl, verhindertsei“ in der Berliner Zeitung publiziert wurde:¹³

Wenn durch diese Entwicklung in ferner Zukunft bald niemand mehr korrektes Deutsch verstehen kann, wird Heinrich Heines „Deutschland, ein Wintermärchen“ im Unterricht bald so vermittelt werden: „an Deutschland in der Nacht denk, nicht kann schlafen“.

Die Schüler würden dann vielleicht „lol“ sagen. In Aufsätzen würden sie „zwinker“ hinter zweideutige Sätze schreiben und der Lehrer würde darunter ein „Beeindrucktsei“ vermerken.¹⁴

Weil diese Elemente in vielen Bereichen als typisch für schriftsprachliches Handeln im Internet angesehen werden, haben wir sie in den Mittelpunkt unserer korpusbasierten Untersuchungen zur Sprach- und Stilvariation gestellt. Gegenüber anderen Merkmalen der internetbasierten Kommunikation – unklare Ganzsatzgrenzen, Abweichungen von orthographischen Normen etc. – können diese Elemente aufgrund ihrer klaren Formeigenschaften relativ gut automatisch ausgezählt werden. Sie eignen sich deshalb auch für Vergleichsstudien in sehr großen Korpora. Für die Beschreibung der Untersuchungen und der Interpretation der Ergebnisse erwies es sich als hilfreich, einen griffigen Ausdruck für diese Elemente zu haben; wir haben uns für den Ausdruck „Stilmarker“ entschieden.

3. Sprachstil und Sprachvariation im Dortmunder Chat-Korpus

3.1 Motivation und Fragestellungen

Chat ist – neben dem Instant Messaging – der Prototyp einer interaktionsorientierten Schreibtechnologie: Im Gegensatz zu Wikipedia-Diskussionen, E-Mails, Diskussionsforen oder der Microblogging-Plattform Twitter ist die Chattechnologie vom Konzept her auf ein kommunikatives Setting angelegt, in dem die Akteure gleichzeitig online sind, das aktuelle Kommunikationsgeschehen direkt verfolgen und mit eigenen Beiträgen darauf reagieren. Das Beispiel in Abbildung 1 illustriert, dass insbesondere die in Flirt- und Plauderchats produzierten Beiträge stark vom Duktus der gesprochenen Sprache geprägt sind und Normen zur Interpunktion und Substantivgroßschreibung weitgehend ignoriert werden. Weil die Beiträge in der Reihenfolge angezeigt werden, in der sie beim Chat-Server eintreffen, stehen inhaltlich aufeinander

¹³ Der von Andreas Weiland verfasste Beitrag ist am 4. Juli 2005 in der Rubrik „Jugend und Schule“ erschienen: <http://www.berlinonline.de/berliner-zeitung/archiv/.bin/dump.fcgi/2005/0704/jugendundschule/0043/index.html>.

¹⁴ Dass die Strophen „Denk ich an Deutschland in der Nacht, dann bin ich um den Schlaf gebracht“, auf die der Autor vermutlich anspielt, aus Heines „Nachtgedanken“ und nicht aus „Deutschland. Ein Wintermärchen“ stammen, ist für unseren Zusammenhang unerheblich.

bezogene Beiträge nicht unbedingt hintereinander. In Abbildung 1 sind beispielsweise die Beiträge zum durchlaufenden Thema „Heiraten – pro und kontra“ durch Sequenzen zur Begrüßung und Verabschiedung von Chattern unterbrochen. Um die Rekonstruktion der Bezüge zu erleichtern, werden Beiträge oft explizit adressiert. In Abbildung 1 adressiert z.B. der Chatter *stoeps* seinen Beitrag 1090 an den Chatter *Dussel* und macht damit klar, dass „lol“ als Rückmeldung auf Dussels „Heiratsantrag“ in Beitrag 1087 zu verstehen ist.

1079	madcow2	nee aber ich hör dauernd wann wollen wir uns denn verloben? willst du mcih überhaupt noch?
1080	madcow2	*blabla blubb*
1081	stoeps	ups
1082		<i>Dussel hört sowas nie</i>
1083	system	Topper betritt den Raum.
1084	Topper	Hi Leute!
1085	stoeps	hi topper
1086	TomcatMJ	der haken is nur,wenns in die brüche geht dann wirts eventuell verdammt teuer...scheidungsanwälte sind nich gerade die günstigsten....
1087		Dussel fragt mal bei stoeps an: wollen wir heiraten? polygamie ist wieder im kommen...
1088		<i>luckyluke verdrückt sich dann ma wieder</i>
1089	luckyluke	bye ihr leuts
1090	stoeps	*lol* dussel
1091	TomcatMJ	by eluke
1092	stoeps	bye luke
1093	luckyluke	frohes feiern
1094	TomcatMJ	tach topf
1095	madcow2	meine kleine hat schon n ehevertrag fertig *ggg*

Abbildung 1: Mitschnitt aus einem Plauderchat (Quelle: Dortmund Chat-Korpus¹⁵)

Chats gehören – gerade wegen ihrer sprachlichen und strukturellen Auffälligkeiten – mittlerweile zu den am besten erforschten Formen der internetbasierten Kommunikation. Sowohl für die sprachlichen Besonderheiten als auch für die im Vergleich zu mündlichen Gesprächen merkwürdigen Strukturen liegen linguistisch fundierte, an Beispielkorpora belegte Erklärungsansätze vor¹⁶. In der interdisziplinären Chat-Forschung wurden Chats in verschiedenen sozialen Kontexten – Beruf, Bildung, Medien – untersucht. Dabei zeigte sich, dass der Sprachstil vom sozialen und institutionellen Setting mindestens ebenso stark geprägt ist wie von den technischen Rahmenbedingungen der jeweiligen Chatumgebung¹⁷. Einen ersten Eindruck von der großen Variationsbreite erhält man, wenn man Abbildung 1 mit dem in Abbildung 2 gezeigten Beispiel vergleicht, das in einem Chat-Angebot der Universitätsbibliothek der TU Dortmund mitgeschnitten wurde. In diesem Mitschnitt siezen sich die Beteiligten, sie verzichten auf Emotikons oder Aktionswörter und orientieren sich sehr viel stärker an orthographischen Normen. Dies liegt sicherlich daran, dass die Interaktion in einem institutionellen Kontext verankert ist, dessen Rollenkonstellation (UB-Mitarbeiter – UB-Nutzer) und die daran gebundenen Verhaltenskonventionen übernommen werden. Ein weiterer Unterschied zum Mitschnitt in Abbildung 1 besteht darin, dass die Chatter einen Zwei-Personen-Dialog führen, in dessen Rahmen sie ihre Beiträge ohne Eile produzieren und sich ohne Unterbrechung durch andere Chatter austauschen können – an den Zeitmarken wird ersichtlich, dass sich die Inter-

¹⁵ Dokument Nr. 2221008; online unter http://www.chatkorpus.tu-dortmund.de/files/releasehtml/htmlkorpus/unicum_03-03-2003.html.

¹⁶ Vgl. die Beiträge im Sammelband Beißwenger (2001a) sowie die Monographien Beißwenger (2000) und (2007), Thaler (2003, zum französischen Chat), Dorta (2005) und Luckhardt (2009).

¹⁷ Vgl. Beißwenger und Storrer (2005) und Storrer (2007).

aktion über insgesamt 5 Minuten hinzieht. Bereits der Vergleich der beiden Mitschnitte macht also die in der Linguistik vertretene Auffassung plausibel, dass nicht die Kommunikationsform des Chats als solche den sprachlichen Duktus bestimmt, sondern dass verschiedene mediale, kontextuelle und soziale Parameter den Sprachstil beeinflussen.

[11:57:16] **BENUTZER** Können Sie mir sagen, ob das Buch Fn 25665 ausgeliehen ist? ¹⁸
 11:51:26 **AUSKUNFT** Hallo, wenn Sie einen Moment Geduld haben, schaue ich im Regal nach - Moment
 11:51:39 **BENUTZER** danke
 11:53:33 **AUSKUNFT** Ist da, ich lege es Ihnen bei der Information im Erdgeschoss zurück, wenn Sie mir bitte Ihren Namen schreiben.
 11:54:11 **BENUTZER** Melanie Treitinger - bis wann muß ich es abgeholt haben?
 11:54:43 **AUSKUNFT** Bis wann schaffen Sie es, dann mache ich den entsprechenden Hinweis dran?
 11:55:19 **BENUTZER** heute oder morgen
 11:55:30 **AUSKUNFT** O.k. dann schreibe ich bis morgen drauf.
 11:55:41 **BENUTZER** Vielen Dank!
 11:56:00 **AUSKUNFT** Gern geschehen und schönen Tag noch.
 11:56:09 **BENUTZER** Ebenfalls
 11:56:16 **system** *** BENUTZER hat den chat verlassen. ***

Abbildung 2: Mitschnitt aus dem UB-Chat der TU Dortmund (Quelle: Dortmunder Chat-Korpus¹⁹)

Um die Zusammenhänge zwischen Sprach- und Diskursstrukturen und den verschiedenen technischen, kontextuellen und sozialen Parametern in der Chatkommunikation auch quantitativ untersuchen zu können, wurde an der TU Dortmund ein Korpus mit Mitschnitten aus verschiedenen Handlungsbereichen – Freizeit, Bildung, Medien, Beruf – aufgebaut und texttechnologisch aufbereitet (Details s.u. 3.2). Auf dieser Ressource – dem Dortmunder Chat-Korpus – basieren die im Folgenden dargestellten Studien.

3.2 Datenbasis: *Das Dortmunder Chat-Korpus*

Das Dortmunder Chat-Korpus²⁰ wurde in den Jahren 2002–2008 aufgebaut und umfasst ca. 140.000 Nutzerbeiträge bzw. 1,06 Millionen Tokens aus unterschiedlichen sozialen Handlungsbereichen: „Plauderchats“ aus dem Freizeitbereich (IRC und Webchats), Chatmitschnitte aus den Bereichen Bildung (E-Learning, studentische Beratung), Chats in medialen Kontexten (Flankierend zu TV-Sendungen, Politchats) und Mitschnitte aus der beruflichen Chatnutzung (Beratung, Service). Alle Mitschnitte wurden in mehreren semiautomatisch und manuell durchgeführten Aufbereitungsschritten in ein XML-Format überführt, das die Struktur der Mitschnitte und der Komponenten der Beiträge (Nickname, ggf. Timestamp, Beitragstext) explizit erfasst. Weiterhin wurde annotiert, ob es sich um Systemmeldungen (z.B. Beitrag 1083 in Abbildung 1) oder um von menschlichen Nutzern generierte Beiträge handelt. Auf dieser Basis lassen sich Elemente wie Nicknames und Systemmeldungen, die in „Bag-of-

¹⁸ In der vom UB-Chat verwendeten Software wird die Verbindung mit dem UB-Mitarbeiter erst nach erfolgter Anliegensformulierung hergestellt. Die Anliegensformulierung wird als erster Chat-Beitrag des Benutzers angezeigt. Bei der automatischen Archivierung des Chat-Protokolls ordnet die Software diesem ersten Benutzerbeitrag, der als Formulierung des thematischen Rahmens für die Chat-Sitzung aufgefasst werden kann, nachträglich den Zeitpunkt der Archivierung der Chat-Sitzung als Timestamp zu. Aus diesem (technischen) Grund ist im hier wiedergegebenen Beispiel dem ersten Benutzerbeitrag ein späterer Timestamp zugeordnet als dem letzten, obwohl er den Auftakt der Chat-Sitzung bildete.

¹⁹ Dokument Nr. 1203043; online unter http://www.chatkorpus.tu-dortmund.de/files/releasehtml/htmlkorpus/UB_Gesamt.html.

²⁰ Vgl. <http://www.chatkorpus.tu-dortmund.de/>.

words“-Ansätzen zur statistischen Auswertung von Chatdaten die Ergebnisse kontaminieren, bei den quantitativen Analysen (s.u.) ausfiltern.

Bei den nutzergenerierten Beiträgen sind die sog. Aktionsbeiträge („action messages“) explizit gekennzeichnet, in denen Chatter über sich in der dritten Person schreiben (z.B. die Beiträge 1087 und 1088 in Abbildung 1). Weiterhin sind typische Stilmarker der internetbasierten Kommunikation annotiert, u.a. die Emotikons und die Aktionswörter (asterisk expressions)²¹. Ebenfalls annotiert wurden explizite Adressierungen (z.B. in den Beiträgen 1090, 1091, 1092 in Abbildung 1); auf diese Weise kann ausgewertet werden, welche Akteure in einer Community besonders häufig angesprochen werden und welche besonders intensiv miteinander kommunizieren²². Für sämtliche Chatter sind Metadaten erfasst, u.a. das vermutete Geschlecht sowie die Anzahl der von ihnen produzierten Beiträge und Tokens.²³

Für die Auswertung des Korpus und der darin verfügbaren Annotationen steht das spezialisierte Werkzeug STACCADo zur Verfügung.²⁴ Mit diesem Werkzeug lassen sich Suchanfragen formulieren oder statistische Auswertungen zu einzelnen Chattern, Chatmitschnitten oder Teilkorpora erzeugen. Die im Korpus annotierten Stilmarker können gefiltert, gezählt und mit frei wählbaren Kontexten ausgegeben werden. Weiterhin erzeugt STACCADo sog. „Logfile-Profile“, in denen die durchschnittlichen Längen der nutzergenerierten Beiträge und die Vorkommensfrequenz von Stilmarkern – Emotikons, Aktionswörtern, Aktionsbeiträgen – zu Chat-Mitschnitten erfasst sind. Solche Logfile-Profile bilden die Basis für die Untersuchungen zur Stilvariation in Abschnitt 3.3. Die „Chatter-Profile“ liefern Übersichten zu den Anteilen einzelner Chatter am Beitragsaufkommen in einem Teilkorpus und ermitteln die Frequenz der verwendeten Stilmarker. In der in 3.4. diskutierten Studie wurden Chatter-Profile u.a. für die Auswahl der untersuchten Akteure in der Community (die Ü600-Gruppe mit insgesamt mehr als 600 Beiträgen) und zur quantitativen Analyse individueller Stilpräferenzen genutzt.

3.3 Sprachvariation in Abhängigkeit von Handlungsbereichen und Kommunikationsphasen

In der sprachwissenschaftlichen Forschung wird immer wieder darauf hingewiesen, dass die sprachliche Form von Chat-Beiträgen in Chats nicht an die Kommunikationsform des Chats und die Funktionen der zugrundeliegenden Kommunikationstechnologie gebunden ist, sondern an das kommunikative Setting, in dem das Kommunikationsereignis stattfindet. Bereits Runkehl et al. (1998: 81) stellten in Bezug auf den Unterschied zwischen moderierten und unmoderierten Chats fest, dass „es sehr unterschiedliche Chats gibt“ und dass deshalb „Pauschalaussagen über das Chatten, wie man sie all zu häufig findet, problematisch sind“. Beim Zusammenstellen des Dortmunder Chat-Korpus haben wir deshalb darauf geachtet, Mitschnitte aus einer möglichst großen Bandbreite an kommunikativen Handlungsbereichen zu erfassen; es finden sich Beispiele aus dem „traditionellen“ Internet Relay Chat (IRC) als auch aus

²¹ Bei der semi-automatischen Annotation wurden allerdings nur die mit „*“ markierten Aktionswörter explizit gekennzeichnet, deshalb der Ausdruck „asterisk expressions“.

²² Eine Analyse von Adressierungsstrukturen in einer Chat-Community findet sich bei Luckhardt (2009: 148ff.).

²³ Zur Aufbereitung der Daten für das Dortmunder Chat-Korpus und zum Aufbau des XML-Formats vgl. die Dokumentation unter <http://www.chatkorpus.tu-dortmund.de/files/korpusaufbereitung.pdf>.

²⁴ STACCADo (*Search Tool for Annotated Chat Corpus Analyses Dortmund*) ist eine GUI-basierte Java-Anwendung, die im Rahmen des Projekts von Bianca Selzam programmiert wurde und kostenfrei auf der Website zum Download bereitsteht. Ebenfalls frei verfügbar, teilweise auch direkt als HTML-Dateien einsehbar, sind Teile des Korpus, in denen in einigen Dokumenten aber Teilnehmernamen sowie Hinweise auf Personen und geographische Orte anonymisiert wurden.

verschiedenen Typen von unmoderierten und moderierten Webchats. Gemeinsames Merkmal ist lediglich, dass sie mit schriftbasierten Standard-Chat-Systemen erzeugt wurden, d.h. Grafik-Chats oder Audio-Chats sind nicht erfasst²⁵. Auf dieser Basis kann man nun Stichproben von Mitschnitten aus unterschiedlichen sozialen Handlungsbereichen zusammenstellen und mit STACCADo auswerten. Unter dem Namen *BalaCK* („Balanced Chat-Korpora“) gibt es bereits vier „vorgefertigte“ Stichprobenkorpora aus den Beständen des Dortmunder Chat-Korpus, die Mitschnitte aus den vier Bereichen Freizeit (sog. „Plauder“-Chats), E-Learning (Seminar-Chats), Beratung und Medien (Polit-Chats etc.) enthalten, wobei in den vier Bereichen entweder die Zahl der Textwörter (Tokens) oder die Zahl der Beiträge (Messages) übereinstimmt²⁶.

Teilkorpus	Beiträge nutzer- generiert	Tokens nutzer- generiert	Durch- schnittl. Anzahl Tokens pro Beitrag	Asterisk Expressions		Emotikons	
				Vorkommen absolut	<i>Relative Frequenz</i>	Vorkommen absolut	<i>Relative Frequenz</i>
Freizeit	6000	25786	4,30	1077	0,042	551	0,021
E-Learning	6000	44980	7,50	42	9,337E-4	320	0,007
Beratung	6000	71037	11,84	10	1,408E-4	189	0,003
Medien	6000	78142	13,02	110	0,001	124	0,002

Tabelle 1: Auswertung der Logfile-Profile für das Vergleichskorpus *BalaCK 2b*.

Unsere quantitative Vergleichsstudie, deren Ergebnisse in Tabelle 1 zusammengefasst sind, basiert auf dem Stichproben-Korpus *BalaCK 2b*, das Stichproben mit je 6000 Beiträgen aus allen vier Handlungsbereichen enthält. Maßgeblich war die Zahl der nutzergenerierten Beiträge (inkl. action messages), da systemgenerierte Beiträge immer wieder dieselben Textbausteine – z.B. „betritt den Raum“, „verlässt den Raum“ – enthalten und damit statistische Auswertungen schnell verzerren können. Auf der Basis der mit STACCADo erstellten Logfile-Profile wurden für alle vier Bereiche die folgenden Merkmale ausgewertet: Die durchschnittliche Länge der Beiträge, gemessen an der Gesamtzahl der von den Nutzern generierten Tokens²⁷; die Anzahl der Vorkommen von Emotikons und die Anzahl der Vorkommen von Aktionswörtern. Die Anzahl der Aktionsbeiträge wurde nicht verglichen, weil viele Mitschnitte vom Typ Seminarchats, Politchats und Beratungschats aus Chatumgebungen stammen, die keine Funktionen zum Erzeugen von Aktionsbeiträgen anbieten.

Die Ergebnisse dieser Auswertung in Tabelle 1 machen deutlich, dass sich die beim Vergleich von Abbildung 1 und 2 beobachtbaren Unterschiede quantitativ nachweisen lassen. Die Mitschnitte aus den sog. Plauderchats sind durchschnittlich wesentlich kürzer und enthalten anteilig sehr viel mehr Emotikons und Aktionsbeiträge als die Mitschnitte in den Vergleichskorpora.

²⁵ Einen Überblick über Chatsysteme gibt Beißwenger (2007: 56ff./Kap 3.3.); dort wird auch der Ausdruck „Standard-Chat-System“ definiert (S. 64).

²⁶ Zur Zusammensetzung im Detail vgl. <http://www.chatkorpus.tu-dortmund.de/korpora.html#balack>.

²⁷ Weil sowohl die interne Struktur der Beiträge als auch die Beitragstypen im Korpus explizit annotiert sind, können bei der Berechnung sowohl die Systembeiträge als auch die ebenfalls automatisch vom System hinzugefügten Nicknames zu Nutzerbeiträgen aus der Berechnung ausgeklammert werden.

Für den Befund, dass in Plauderchats sehr viel kürzere Beiträge geschrieben werden als in den anderen Handlungsbereichen, bietet die Linguistik plausible Erklärungsmuster an. Z.B. hat bereits Beißwenger (2000: 45 ff.) darauf hingewiesen, dass im kommunikativen Setting eines Plauderchats viele Beiträge dazu dienen, fortbestehende Aufmerksamkeit und Kommunikationsbereitschaft zu signalisieren, hier genügen sehr kurze Rückmeldungen. Auch bei den Begrüßungs- und Verabschiedungsritualen entstehen – wie in Abbildung 1 sichtbar – viele kurze Beiträge. Beim interaktiven Schreiben in Mehrpersonenchats, bei denen Beiträge in schneller Abfolge über den Bildschirm laufen, ist es vor allem wichtig, kommunikativ „am Ball“ zu bleiben, auch hierfür sind mehrere kurze Beiträge funktional angemessen: Da die Beiträge für die Chatpartner erst sichtbar werden, wenn sie am Bildschirm erscheinen, unterbricht das Formulieren längerer Beiträge den dialogischen Kommunikationsfluss. Gerade im Kontext des E-Learning hat sich die Technik des sog. „Splitting“ bewährt, bei dem Produzenten ihre Beiträge in mehrere hintereinander geschaltete Chunks aufbrechen und damit einen kontinuierlicheren Lesefluss ermöglichen.²⁸ Dies könnte auch erklären, warum die durchschnittliche Länge der Beiträge in Seminarchats zwar größer ist als die der Plauderchats, aber doch geringer als man dies bei Seminardiskussionen erwarten würde.

38	Betzebub	Der Countdown läuft
39	Humanist	jo
40	[zdf]axel	Richtig.
41	Erik89	smakelijk eten hahahaha
42	internetzer2004	loddar wir lieben dich
43	Der Bomber	Hallo!! Na schon gespannt auf heute Abend???????????
44	zeneca	weil di halt defiziten haben
45	bart2004	weil die vielleicht echt schlecht sind.
46	Betzebub	Wir habens drauf;)
47	DeutschlandFan161616	genau, nicht normal
48	[zdf]aring	Wenn Lothar Matthäus im Chat ist, wird der Chat moderiert - das heißt die Fragen werden gesammelt und einzeln freigestellt
49	PassionMan85	Wer teilt meine meinung,das Deutschland 3:1 gewinnt?
50	internetzer2004	ich
51	aligo	Es kam über die Jahre und jetzt sitzt es ziemlich fest.Solangs um Fußball geht... ;-)
52	Erfurter	ich
53	Wolf2362	ich auch
54	internetzer2004	3:1
55	Humanist	ich

Abbildung 3: Mitschnitt aus der Vorlaufphase eines „Promichats“²⁹

Dass im Bereich der medialen Chats relativ lange Beiträge produziert werden, lässt sich ebenfalls mit dem technischen und kommunikativen Setting der hier untersuchten Mitschnitte erklären. Es handelt sich größtenteils um moderierte Chats, in denen Moderatoren darüber entscheiden, welche der von den Chattern gestellten Fragen an die Interviewpartner zur Beantwortung weitergereicht werden. In diesem Setting ist also nicht der Zeitpunkt des Eintreffens beim Server, sondern maßgeblich auch die sprachliche und inhaltliche Qualität des Beitrags entscheidend für das Erreichen des kommunikativen Handlungsziels (= die Auswahl der Frage durch den Moderator und die nachfolgende Beantwortung durch den Interviewgast). Des-

²⁸ Vgl. dazu die Beispiele in Lemnitzer und Naumann (2001); Storrer (2001: 454ff.); Beißwenger (2007: 245-253).

²⁹ Dortmunder Chat-Korpus, Dokument Nr. 1305002; Interview mit Lothar Matthäus, zum Zeitpunkt des Chats (15.6.2004) Bundestrainer der ungarischen Nationalmannschaft; online unter http://www.chatkorpus.tu-dortmund.de/files/releasehtml/html-korpus/zdf_Lothar_Matthaeus_15-06-2004.html.

halb bleibt wieder mehr Zeit für die Sprachplanung und die Formulierung längerer, sprachlich elaborierter Beiträge. Interessant für unseren Untersuchungszusammenhang ist, dass es in vielen moderierten Chats einen unmoderierten Vorlauf und Ausklang gibt, in dem die Chatter vor dem Eintreffen des Interviewpartners und nach dessen Verabschiedung ohne moderierenden Eingriff miteinander kommunizieren können. Bereits Diekmannshenke (2005: 135ff.) hat für Politchats gezeigt, dass in diesen Phasen wieder der für das interaktionsorientierte Schreiben typische Sprachduktus vorherrscht. Der Vergleich der Mitschnitte in Abbildung 3 (Vorlaufphase) und 4 (Interviewphase) illustriert diese Unterschiede in Struktur und sprachlichem Duktus; beide stammen aus einem Chat mit Lothar Matthäus, der am 15.6.2004 wenige Stunden vor der Partie Deutschland gegen die Niederlande während der Fußball-Europameisterschaft in Portugal mitgeschnitten wurde.

- 284 **Phantom+** Sehr geehrter Herr Matthäus, glauben sie, dass Michael Ballack die in ihn gesetzten Erwartungen erfüllen kann?
- 285 **Lothar Matthäus** Ich hoffe es.
- 286 **Alter Preuße** Philip Lahm erinnert mich in seiner Dynamik, perfekten Ballbehandlung und Kraft sehr an den jungen Lothar Matthäus. Stimmen Sie mir da zu, Herr Matthäus?!
- 287 **Lothar Matthäus** Er spielt auf einer anderen Position, aber er hat zumindest den gleichen Charakter wie ich in jungen Jahren.
- 288 **Humanist** Herr Matthäus, glauben Sie an einen EM-Sieg der Deutschen oder halten sie dass für abwegig?
- 289 **the duke** Was glauben sie wie weit Deutschland kommt?
- 290 **Lothar Matthäus** Im Fußball ist alles möglich.
- 291 **maxx18** Herr Matthäus...wenn Sie einen Spieler nachnominieren könnten...wer wäre dies ? Ein junger Wilder wie Benjamin Auer, ein verärgerter Jörg Böhme und jemand ganz anderes ?
- 292 **Lothar Matthäus** Kommt auf die Position an.
- 293 **tim0r77** Herr Matthäus; Wird Bobic von Anfang an spielen oder rechnen sie mit Podlski und Kurani ?
- 294 **Lothar Matthäus** Ich rechne da nur mit Kuranyi.
- 295 **holgi84** Was meinen sie Herr Matthäus, ist es nicht ein wenig übertrieben, wie Herr Völler und Michael Skibbe die Taktik der Deutschen Mannschaft verheimlichen? Bzw. was ist der Sinn der Sache, es ist doch eigentlich fast alles klar wie gespielt wird.
- 296 **Lothar Matthäus** Man versucht, den Gegner noch zu verunsichern, um nicht alle Karten aufzudecken.
- 297 **EhEhm** Glauben Sie Podolski oder Schweinsteiger könnten heute eine Jokerrolle spielen oder wird Rudi sie bei einem so wichtigen SPIEL nicht einwechseln ?
- 298 **Lothar Matthäus** Beides ist möglich. Wenn nötig, wird Rudi sie bringen.

Abbildung 4: Mitschnitt aus der Interviewphase desselben Chats wie in Abb. 3

Bislang sind im Korpus die unterschiedlichen Phasen in medialen Chats nicht explizit annotiert, sodass man nicht automatisch auswerten kann, wie sich die Emotikons und Aktionswörter über die Phasen hinweg verteilen. Stichproben zeigen, dass sehr viele davon aus den unmoderierten Chatphasen stammen. Der Vergleich von Beiträgen desselben Chatters in verschiedenen Phasen – vgl. die Beispiele in Abbildung 5 – kann Indizien dafür liefern, dass sich auch beim interaktionsorientierten Schreiben mehrere Stilregister ausbilden, aus denen die Akteure diejenigen auswählen, die aus ihrer Sicht zur jeweiligen Gesprächsphase am besten passen. Dies spricht für die Auffassung, dass die Kompetenz zum situationsangemessenen Schreiben durch die neuen interaktionsorientierten Schreibformen nicht verloren geht, sondern sich lediglich in neuer Weise ausdifferenziert. Diese in Abschnitt 2 als „Hypothese des

sprachlichen Ausbaus“ eingeführte Sichtweise wird durch die Ergebnisse der in Abschnitt 4 diskutierten quantitativen Vergleichsstudie weiter gestärkt.

Moderierte Phase:

57 eCoach Es wird behauptet, Kernenergie sei sauber. Wieviel Schaden entsteht denn nun durch Uranbergbau und Freisetzung radioaktiver Edelgase - Ist Kernenergie "fast" sauber oder "Fast so dreckig wie fossile Brennstoffe ?

Ausklangphase

247 eCoach @wilhelmyne: Die violetten Blödeln rum, die blauen versteh ich nicht und die schwarzen halten sich nicht an die Farbtrennung. Wilhelmyne, lass uns doch auf GRÜN schalten , damit man die sinnvollen Beiträge sofort erkennt

409 eCoach hauptsache, der wird jetzt auch benutzt. Was fährste nu ?

452 eCoach LOL

Abbildung 5: Beiträge desselben Chatters in verschiedenen Phasen eines Experten-Chats³⁰

3.4 Individuelle Sprachvariation und stilistische Präferenzen

Eine umfangreiche korpuslinguistische Studie zum Sprachstil in einer Chat-Community hat Kristin Luckhardt in ihrer Dissertation vorgelegt (Luckhardt 2009). Die Arbeit verbindet Konzepte und Methoden zur Stilanalyse von Barbara Sandig (1986, 2006) mit korpusgestützten Methoden, um grundlegende stilistische Handlungsmuster – Ökonomisieren, Emotionalisieren, Solidarisieren – und die dafür typischen stilistischen Verfahren herauszuarbeiten. Datenbasis sind Mitschnitte aus einer Chat-Umgebung, die als Teil eines medialen Portals Informationen rund um das Radsportereignis „Tour de France“ bereithält. Das Korpus umfasst Mitschnitte aus dem Zeitraum vom 5.–27.7. 2003; die Chatter waren über das Forschungsvorhaben informiert. 55 Mitschnitte wurden mit den für das Dortmunder Chat-Korpus genutzten Verfahren annotiert und mit der Recherche- und Analysesoftware STACCADo ausgewertet (vgl. Luckhardt 2009: 10ff.). Im Folgenden möchte ich aus den Ergebnissen den Aspekt der individuellen Sprachvariation herausgreifen.

In ihrem korpusgestützten Ansatz arbeitet Kristin Luckhardt die stilistischen Handlungsmuster nicht nur an Beispielen heraus, sondern erhebt zusätzlich Daten zur Häufigkeit typischer Verfahren zur Durchführung dieser Muster. Denn für viele stilistische Handlungsmuster gibt es alternative stilistische Verfahren: So kann z.B. beim Handlungsmuster „Emotionalisieren“ entweder ein Emotikon, ein Aktionswort (z.B. **freu**) oder ein Aktionsbeitrag (z.B. *engelnchen freut sich auf sonntag...*) verwendet werden. Mit Hilfe von STACCADo hat Luckhardt analysiert, wie häufig die Chatter der untersuchten Community von diesen drei Optionen Gebrauch machen. Dabei konzentrierte sie sich auf eine Gruppe von Chattern, die im Untersuchungszeitraum insgesamt mehr als 600 Chat-Beiträge verfasst hatten (die sog. Ü600-Gruppe). In Tabelle 2 sind die Ergebnisse der Frequenzanalysen zu Emotikons, Aktionswörtern und Aktionsbeiträgen für die einzelnen Chatter (alphabetisch nach deren Nicknames ge-

³⁰ „Moderierte Phase“ = Phase mit restriktiver, technisch unterstützter Moderation im Beisein des Experten / „Ausklangphase“ = Phase im Anschluss an die moderierte Phase, in der die Moderation und die technische Regulierung des Beitragsaufkommens wieder aufgehoben sind; die kommunikativen Rahmenbedingungen entsprechen in dieser Phase denen eines unmoderierten Plauderchats (Quelle: Dortmunder Chat-Korpus, Dokument Nr. 1303001; Experten-Chat mit Hans-Joachim Schellnhuber vom Potsdam-Institut für Klimafolgenforschung im Anschluss an eine vorangegangene TV-Diskussion in der ZDF-Sendung *Christiansen* zum Thema „Hochwasser-Katastrophe: Die Flut und die Folgen...“ vom 18.8.2002; online unter http://www.chatkorpus.tu-dortmund.de/files/releasehtml/html-korpus/Sabine_Christiansen_Hans-Joachim_Schellnhuber_userchat_18-08-2002.html).

ordnet) zusammengefasst; einmal in absoluten Zahlen, einmal anteilig in Bezug auf die Gesamtmenge der von der Gruppe verwendeten Stilmarker dieses Typs³¹.

Chatter	Emotikons	Anteil Emotikons	Aktionswörter	Anteil Aktionswörter	Aktionsbeiträge	Anteil Aktionsbeiträge
Benzi 1	23	3,97%	65	15,63%	26	1,06%
Coco	11	1,90%	30	7,21%	18	0,73%
Danidrach	20	3,45%	2	0,48%	0	0%
Engelchen	45	7,76%	73	17,55%	390	15,84%
Florentine	9	1,55%	63	15,14%	123	5,00%
Hernanb	121	20,86%	10	2,40%	41	1,67%
Hexmex	40	6,90%	4	0,96%	9	0,37%
Juergene	1	0,17%	9	2,16%	669	27,17%
Lupus	9	1,55%	2	0,48%	19	0,77%
Ninja 1	52	8,97%	28	6,73%	41	1,67%
Somme ma	150	25,86%	1	0,24%	18	0,73%
Tetedela	5	0,86%	12	2,89%	242	9,83%
tourteam	1	0,17%	0	0%	0	0%
Ursi99	42	7,24%	2	0,48%	0	0%
Uweampller	0	0%	18	4,33%	297	12,06%
Zoe 34	3	0,52%	95	22,84%	18	0,73%
181074	1	0,17%	0	0%	387	15,72%
221262	47	8,10%	2	0,48%	164	6,66%
Insgesamt	580	100%	416	100%	2462	100%

Tabelle 2: Individuelle Stilpräferenzen in einer Chatter-Community

Die Daten illustrieren die große individuelle Variation: Einige Chatter, z.B. *Engelchen* oder *florentine*, machen von allen drei stilistischen Verfahren Gebrauch. Andere Chatter haben klare Präferenzen: *uweampller*, *juergene* und *181074* produzieren sehr viele Aktionsbeiträge, aber kaum Emotikons; *Somme ma* und *Hernanb* haben eine Präferenz für Emotikons; *Zoe34* für Aktionswörter. Interessant ist, dass der Chatter *tourteam* mit einer einzigen Ausnahme keines dieser dem Handlungsmuster „Emotionalisieren“ zugeschriebenen Verfahren nutzt, was Luckhardt (2009: 133f.) plausibel auf dessen Rolle als ein vom Anbieter des Chatangebots bezahlter Radsportexperte zurückführt. Interessant ist auch, dass vor allem die männlichen Chatter gerne und häufig Aktionsbeiträge produzieren, mit denen sich das kommunikative Geschehen im Chatraum von einer „Außenperspektive“ kommentieren und kreativ ausgestalten lässt³², während der häufige Gebrauch von Emotikons eher eine Stilpräferenz weiblicher Chatter zu sein scheint. Diese Indizien für genderspezifische Unterschiede müssten aber natürlich auf einer sehr viel breiteren Datenbasis und mit einem Ansatz untersucht werden, der quantitative Auswertungen mit qualitativen, gesprächsanalytischen Methoden verbindet.

³¹ Die Zahlen sind zusammengestellt aus verschiedenen Aufstellungen in Luckhardt (2009: 110f., 114, 117, 127).

³² Der Stellenwert des Verfahrens für die Konstitution der Gruppe und eines Gruppenstils wird in den qualitativen Analysen zum Handlungsmuster „Solidarisieren“ sichtbar, vgl. die Beispiele in Luckhardt (2009: 176f., 182f.).

Die Arbeit von Luckhardt (2009), aus der hier nur ein Aspekt herausgegriffen wurde, dokumentiert, dass sich qualitative und quantitative, korpusbasierte Analysen sehr gut ergänzen.³³ Einerseits sind Recherchewerkzeuge, mit denen gezielt nach sprachlichen Merkmalen in Korpora gesucht werden kann, sehr nützlich für die Hypothesengenerierung und die explorative Auswertung des Datenbestandes. Andererseits können die Werkzeuge zur statistischen Auswertung – z.B. die Chatter- und Logfile-Profile – dazu beitragen, die an Einzelbeispielen gewonnenen Hypothesen mit quantitativen Methoden an einem größeren Datenbestand zu überprüfen und damit empirisch zu untermauern. Das Potenzial der Kombination von quantitativer und qualitativer Analyse wird zwar bereits am relativ kleinen Dortmunder Chat-Korpus erkennbar. Um im Hinblick auf sprachliche Veränderungen und Sprachvariation aussagekräftige Ergebnisse erzielen zu können, würden aber weitaus größere Datenbestände benötigt.

4. Sprachvariation in der Wikipedia

4.1 Motivation und Fragestellung

In dieser Studie geht es um den Sprachstil und die Sprachvariation in der deutschen Wikipedia. Ausgangspunkt ist die in Abschnitt 2 eingeführte Unterscheidung von textorientiertem und interaktionsorientiertem Schreiben und die Annahme, dass sich die netztypischen Besonderheiten vor allem in den neuen interaktionsorientierten Schreibprozessen herausbilden. Um diese Annahme empirisch zu untermauern, eignen sich kollaborative Wiki-Projekte wie die Wikipedia deshalb gut, weil die für Wiki-Software typische Unterscheidung zwischen Artikel- und Diskussionsseiten (s.u.) sehr gut mit den beiden Schreibhaltungen korrespondiert: Artikeltexte werden eher mit einer textorientierten Schreibhaltung, Diskussionsseiten eher mit einer interaktiven Schreibhaltung verfasst (vgl. auch Diewald und Frank-Job in diesem Band). Da viele Akteure der Wikipedia sowohl an Artikel- als auch an Diskussionsseiten beteiligt sind, kann ein Vergleich der beiden Seitentypen Indizien dafür liefern, dass kompetente Schreiber zwischen den beiden Schreibhaltungen und den jeweiligen Stilen wechseln können. Den ARD/ZDF-Online-Studien 2010/2011³⁴ zufolge ist die Wikipedia das am häufigsten genutzte Angebot des Web 2.0; allerdings ist nur ein sehr kleiner Teil der Nutzer auch selbst aktiv am Aufbau der Wikipedia beteiligt. Da auch nicht alle Nutzer mit dem Unterschied zwischen Artikel- und Diskussionsseiten vertraut sind, möchte ich im folgenden Abschnitt kurz die für die Untersuchung relevanten linguistischen Aspekte des Vergleichs erläutern.

4.2 Artikelseiten vs. Diskussionsseiten in der Wikipedia

Die Wikipedia basiert technisch auf der Software *MediaWiki*³⁵, die – wie andere wikibasierte Systeme – spezielle Funktionen für die gemeinschaftliche Textproduktion bereitstellt. Eine dieser Besonderheiten besteht darin, dass jede Artikelseite mit einer dazugehörigen Diskussionsseite verknüpft ist, auf der über Inhalte der jeweiligen Artikelseite diskutiert werden kann³⁶. Die beiden Seitentypen unterscheiden sich grundlegend sowohl im Aufbau als auch im sprachlichen Duktus; zur Illustration habe ich Ausschnitte aus der Artikelseite zu Ernst

³³ Ein weiteres Beispiel für die Verbindung quantitativer und qualitativer Analysen bei der Arbeit mit dem Dortmunder Chat-Korpus ist die Untersuchung von Michael Beißwenger (in diesem Band) zur Verwendung von Lokaldeiktika in der Chat-Kommunikation.

³⁴ <http://www.ard-zdf-onlinestudie.de/>.

³⁵ <http://www.mediawiki.org/>.

³⁶ Die Diskussionsseiten sind nicht der einzige, aber doch der wichtigste Ort, an dem über die Inhalte der Artikelseiten diskutiert wird; vgl. Stegbauer (2009: 71f.).

Jandls Gedicht „Ottos Mops“ (Abbildung 6) und der dazugehörigen Diskussionsseite abgebildet (Abbildung 7)³⁷.

The screenshot shows the Wikipedia article for "ottos mops". The page layout includes the Wikipedia logo and navigation menu on the left, and the article content on the right. The article text describes the poem by Ernst Jandl, mentioning its date (November 20, 1963) and its publication in the book "der künstliche baum". It also includes a photograph of a pug dog, captioned "Ein Mops".

Abbildung 6: Wikipedia-Artikelseite (*ottos mops*).

The screenshot shows the discussion page for "ottos mops". The page title is "Diskussion:Ottos mops". It features a table of contents with seven entries, including "Review im März/April 2010". The main content of the discussion is a review from March/April 2010, which discusses the poem and the article's content. The review text is as follows:

Das Gedicht über den trotzenden und kotzenden Mops von Ernst Jandl kennen wahrscheinlich die meisten. Eingehendere Untersuchungen gibt es aber kaum, so dass der Artikel recht umfassend sein sollte. Trotzdem würden mich weitere Querverweise interessieren: Wo ist das Gedicht noch zur Aufführung/Anwendung gekommen? Gibt es Kritiken zum Gedicht, vielleicht gar einen Verriss? Vielleicht gräbt ja jemand noch was aus. Es sind aber auch alle sonstigen Anmerkungen, Vorschläge, Verbesserungen im Hinblick auf eine Kandidatur erwünscht. Viele Grüße! --Magiers 00:42, 15. Mär. 2010 (CET)

Gebe Ernsts freche Rede jetzt der Bewerber-Menge eben wegen perfekten Themen-Kennens des Erstellers, ferner wegen des Erstellers Verstehens des Regel-Gemenges der Lese-Perlen. M.E. Exzellente! --Krechz 03:08, 15. Mär. 2010 (CET)

reinquetsch: Wow! Klasse auf -e- :-)) --Felistoria 00:18, 23. Mär. 2010 (CET)

Ich hasse Gedichte... aber nicht diesen wunderbar mopsigen Artikel. Keine Verbesserungsvorschläge. --ly 12:07, 15. Mär. 2010 (CET)

Abbildung 7: Wikipedia-Diskussionsseite (*ottos mops*).

³⁷ Die Ausschnitte wurden am 15.8.2010 fotografiert.

Schon die kleinen Ausschnitte machen zwei zentrale Unterschiede deutlich: Die Artikelseite ist monologisch strukturiert, die Abfolge und Anordnung der Teilthemen folgt Konventionen für den behandelten Thementyp „Gedicht“. Die Normen der schriftlichen Standardsprache sind eingehalten, das sprachliche Produkt entspricht den Erwartungen, die man sprachstilistisch an einen enzyklopädischen Artikel stellt, auch wenn natürlich durch die gemeinschaftliche Arbeit am Text und dessen Hypertextualität (Verlinkung, Veränderbarkeit etc.) Aspekte hinzukommen, die bei der individuellen produktbezogenen Textproduktion fehlen. Die Diskussionsseite hingegen ist chronologisch nach Themen geordnet; innerhalb eines Diskussthemas findet man die für Diskussionsforen typischen „Thread-Strukturen“, d.h. es gibt einen Anfangsbeitrag und darauf bezogene Repliken, die dann (in einem rekursiven Prozess) ihrerseits wieder Gegenstand von Repliken werden können. Gerade die Folgebeiträge zeigen oft die typischen sprachlichen Merkmale der interaktionsorientierten Schreibhaltung, die sich am Duktus der informellen mündlichen Kommunikation orientiert (die englische Bezeichnung „talk page“ für diesen Seitentyp macht diese Orientierung besonders gut deutlich).

Die sprachstilistischen Unterschiede zwischen Artikel- und Diskussionsseiten machen die Wikipedia zu einem interessanten Korpus, an dem die Herausbildung, Entwicklung und individuelle Variation der interaktionsorientierten Schriftlichkeit auch sprachvergleichend untersucht werden kann. Die im Folgenden dargestellte Vergleichsstudie konnte an einem linguistisch nicht annotierten Wikipedia-Korpus nur eine kleine Menge von formal gut erkennbaren Stilmarkern untersuchen. Die Studie zeigt aber das Potenzial der Wikipedia als Datengrundlage für weitere Untersuchungen, wobei auch deutlich wird, warum für weiterführende Untersuchungen eine linguistisch und texttechnologisch aufbereitete Version wünschenswert wäre, die gezielten Zugriff auf die Dokumentstrukturen der Artikelseiten bzw. die Threadstrukturen der Diskussionsseiten (inkl. der entsprechenden Metadaten zu Autoren, Timestamps etc.) erlaubt.

4.3 Datengrundlage und Auswahl der analysierten Stilmarker

In einer Vergleichsstudie untersuchten wir Vorkommensfrequenzen von typischen Stilmarkern in Artikel- und Diskussionsseiten an einer kompletten Kopie der deutschsprachigen Wikipedia vom 18.6.2009, die uns freundlicherweise vom Ubiquitous Knowledge Processing Lab (UKP) der TU Darmstadt bereitgestellt und in eine MySQL-Datenbank importiert wurde. Die Kopie umfasst insgesamt 1.314.853 Seiten, davon sind 1.011.112 Artikelseiten und 303.741 Diskussionsseiten³⁸. Die durch eine Datenbankabfrage ermittelte Gesamtzahl der Tokens beläuft sich auf 692.863.553, davon finden sich 515.593.226 in Artikelseiten und 177.270.327 in Diskussionsseiten. Bei der Datenbankabfrage wurden alle Elemente als Tokens gewertet, die durch zwei Leerzeichen getrennt sind; da die Datenbank die Quelltexte der Seiten erfasst (also nicht die formatierten Versionen, die im Browser angezeigt sind), wurden auch Elemente der HTML-Syntax als Tokens gewertet, wenn diese zwischen zwei Leerzeichen auftraten.

Zur Datenerhebung nutzten wir ein von Bianca Stockrahm entwickeltes Java-Programm namens „WikiSearch“, das sämtliche Artikel- und alle Diskussionsseiten nach einem Suchterm durchsucht und die Treffer unter Angabe des Seitennamens als Liste in eine HTML-Datei schreibt; der Belegschnitt erfasst dabei 50 Zeichen vor und nach dem Suchterm. Die Anbin-

³⁸ Die technische Option, eine Diskussion zu eröffnen, ist prinzipiell immer vorhanden, sie wird aber bei weitem nicht bei jeder Artikelseite genutzt (vgl. auch die Befunde in Stegbauer 2009: 71, 118f.).

dung von „WikiSearch“ an die Datenbank basierte auf der an der TU Darmstadt entwickelten Open-Source-Schnittstelle *JWPL* (Java Wikipedia Library).³⁹

Bei der Zusammenstellung der untersuchten Stilmarker orientierten wir uns an den Vorkommensfrequenzen im Dortmunder Chat-Korpus, in dem die folgenden interaktiven Einheiten am häufigsten belegt sind:

- die Aktionswörter *lol, lach, freu, grins, wink, seufz*; diese wurden im Wikipedia-Korpus sowohl in den formal durch * markierten Varianten als auch in unmarkierter Form ausgewertet;
- die Emotikons :-), :), ;-), ;), :-(), :(

Die Studie beschränkt sich damit auf eine kleine Auswahl von interaktiven Einheiten zur emotionalen Kommentierung, die sich an formalen Kennzeichen relativ gut erkennen und deshalb an einem nicht-annotierten Korpus mit vertretbarem Zeitaufwand untersuchen lassen⁴⁰.

4.4 Datenauswertung und Interpretation der Ergebnisse

Die Ergebnisse der Auswertung der Artikelseiten sind in Tabelle 3, die Ergebnisse der Auswertung der Diskussionsseiten in Tabelle 4 zusammengestellt. Es zeigt sich sehr klar, dass die untersuchten Stilmarker fast ausschließlich auf den Diskussionsseiten vorkommen⁴¹. Auf den Artikelseiten – immerhin 1.011.112 Seiten mit 515.593.226 Tokens – gibt es keinen einzigen Beleg für die Verwendung eines Aktionsworts und nur sechs Verwendungen von Emotikons. Dieser Befund ist in seiner Eindeutigkeit doch überraschend, zumal man davon ausgehen kann, dass die an Wikipedia beteiligten Akteure⁴² mit dem Duktus des interaktionsorientierten Schreibens vertraut sind. Auf den Diskussionsseiten sind die Stilmarker zwar zu finden, allerdings sehr viel seltener als z.B. in der Stichprobe der Plauderchats des Dortmunder Chat-Korpus, wie man am Vergleich der relativen Häufigkeiten in den Tabellen 1 (Abschnitt 3) und 4 sieht.

Dieses eindeutige Ergebnis entsteht allerdings nur, wenn die Trefferliste bereinigt und Pseudotreffer (falsch positive Treffer) ausgesondert werden. Bei den Aktionswörtern, die als unflektierte Verbstämme (Inflektive) gebildet sind, betrifft dies v.a. homonyme Formen der 1. Person Singular (Beispielbeleg 1) und des Imperativ Singular (Beispielbeleg 2).

(1) So viel Spaß für wenig Geld (Ich **freu** mich)⁴³

(2) Karin Jeromin / Rüdiger Pfeffer: Komm **freu** dich mit mir. Die Bibel für Kinder erzählt.⁴⁴

³⁹ Vgl. Zesch, Müller und Gurevych (2008); weitere Informationen zum Tool finden sich unter: <http://www.ukp.tu-darmstadt.de/index.php?id=2634>. Wir bedanken uns bei Torsten Zesch für die tatkräftige Unterstützung.

⁴⁰ Die Auswertung erfolgte im Zusammenhang mit der studentischen Staatsarbeit von Christina Lüken.

⁴¹ Wegen der unterschiedlichen Verteilung der Gesamttokens auf die beiden Seitentypen müssen die absoluten Frequenzzahlen in Relation zur Grundmenge der beiden Seitentypen gesetzt werden; die Angaben zur relativen Frequenz sind deshalb in die Tabellen integriert.

⁴² Vgl. die soziodemographischen Analysen in Stegbauer (2009: 285ff.).

⁴³ Artikelseite „Die Prinzen“.

⁴⁴ Artikelseite „Die Kinderbibel“.

Suchformen	Treffer	Pseudotreffer	Belege	Relative Häufigkeit
lol *lol*	59	59	0	0
seufz *seufz*	0	0	0	0
grins *grins*	46	46	0	0
lach *lach*	170	170	0	0
freu *freu*	47	47	0	0
wink *wink*	266	266	0	0
;-)	8	6	2	3,88E-9
:-)	338	337	1	1,94E-9
:)	127	127	0	0
;))	46	44	2	3,88E-9
:-(36	36	0	0
:(50	49	1	1,94E-9

Tabelle 3: Frequenz der IBK-Stilmarker auf den Artikelseiten.

Suchform	Treffer	Pseudotreffer	Belege	Relative Häufigkeit
lol *lol*	646	11	615	3,47E-6
seufz *seufz*	223	1	222	1,25E-6
grins *grins*	192	19	173	9,76E-7
lach *lach*	180	67	113	6,37E-7
freu *freu*	239	188	51	2,88E-7
wink *wink*	128	123	5	2,82E-8
;-)	25350			1,43E-4
:-)	24211			1,37E-4
:)	10424			5,88E-5
;))	9893			5,58E-5
:-(1438			8,11E-6
:(822			4,64E-6

Tabelle 4: Frequenz der IBK-Stilmarker auf den Wikipedia-Diskussionsseiten.

Fast alle Belege dieses Typs sind Bestandteile von Zitaten oder Titeln von Filmen, Liedern, TV-Sendungen o.ä. Weil das Suchwerkzeug nicht zwischen Groß- und Kleinbuchstaben unterscheiden kann, gab es weiterhin sehr viele Pseudotreffer, in denen homographie Eigennamen ausgegeben wurden; zu diesem Typ gehören sehr viele Pseudotreffer von „lach“ und „lol“ (Beispielbeleg 3).

- (3) 1968 trat der experimentelle Saxophonist **Lol** Coxhill bei, und der Bandname wechselte zu "Steve Miller's Delivery", der übersetzt "Steve Miller's Botschaft" bedeutet.⁴⁵

Bei den Emotikons resultieren viele Pseudotreffer aus formgleichen Zeichenkombinationen, z.B. in zwar bereits angelegten, aber noch nicht ausgefüllten Tabellen zu Sportereignissen wie

⁴⁵ Artikelseite „Delivery (Band)“.

in „Niederlande – Finnland :- (-:-) Samstag, 29. August 2009, 17:30“ auf der damaligen Artikelseite „Fußball-Europameisterschaft der Frauen 2009“.

Weitere Pseudotreffer bei Emotikons haben damit zu tun, dass die Datenbasis die Quelltexte der Seiten erfasst und damit auch Elemente der HTML-Syntax mit entsprechenden Graphemfolgen enthält. Außerdem fanden sich erwartungsgemäß in Wikipedia-Artikeln, die sich mit internetbasierter Kommunikation befassen, auch Treffer wie in Belegbeispiel 4, in denen die betreffenden Ausdrücke als Beispiele erwähnt, aber nicht im Satz verwendet werden.

- (4) Die fehlenden parasprachlichen Mittel werden durch Emoticons (z. B. :-), ;-) oder :-o) und Akronyme (z. B. lol = laugh(ing) out loud; dt. "Lautes Lachen") oder Abkürzungen ersetzt.⁴⁶

Derartige metakommunikative Treffer wurden in den Ergebnistabellen ebenfalls zu den Pseudotreffern gerechnet.

Die Erfahrungen aus der Studie zeigen deutlich, dass eine manuelle Bereinigung der Trefferlisten unerlässlich ist, um ein zuverlässiges Bild von den tatsächlichen Vorkommensfrequenzen zu erhalten: Von den 338 automatisch gezählten Treffern des Emotikons „:-)“ in den Artikelseiten blieb nach der Bereinigung nur ein Beleg übrig. Die Trefferlisten zu allen Aktionswörtern in Artikelseiten enthalten überhaupt nur Pseudotreffer. Bei den Diskussionsseiten liegt der Anteil der Pseudotreffer deutlich niedriger, dennoch lohnt sich die Bereinigung auch deshalb, weil die Anzahl der Pseudotreffer bei den verschiedenen Einheiten erheblich schwankt, sodass es nicht möglich ist, formunabhängige Schätzwerte anzugeben. Bei den Emotikons in den Diskussionsseiten mussten wir aufgrund der hohen Trefferzahlen auf eine manuelle Bereinigung verzichten; die Werte müssen also mit Vorsicht interpretiert werden; allerdings ist der Unterschied zu den (unbereinigten) Vorkommensfrequenzen in den Artikelseiten schon überaus auffällig.

Wegen des hohen Nachbereitungsaufwandes sind vergleichende Untersuchungen zu anderen Sprachmerkmalen, die als typisch für das interaktionsorientierte Schreiben in sozialen Netzwerken gelten, auf linguistisch nicht weiter aufbereiteten Versionen der Wikipedia extrem zeitaufwändig. Da momentan in der Korpuslinguistik an der linguistischen und diskursstrukturellen Aufbereitung der Wikipedia gearbeitet wird, ist zu erwarten, dass bald auch weitere systematischere Studien zur Sprach- und Stilvariation auf Artikel- und Diskussionsseiten durchgeführt werden können. Insgesamt machen die Ergebnisse dieser kleinen Vergleichsstudie deutlich, welches Potenzial sich durch den Vergleich der beiden Seitentypen für die empirische, quantitativ gestützte Untersuchung der sprachstilistischen Unterschiede zwischen interaktions- und textorientierten Schreibformen ergibt.

5 Stilmarker der internetbasierten Kommunikation in der journalistischen Prosa

Die Wikipedia-Vergleichsstudie liefert empirische Indizien dafür, dass die untersuchten Stilmarker tatsächlich an Kontexte des interaktionsorientierten Schreibens gebunden sind. Ergänzend soll nun geprüft werden, ob und wie häufig sie in redigierten Zeitungstexten verwendet werden. Eine solche Untersuchung kann auf der Basis digitaler und linguistisch aufbereiteter Zeitungskorpora heutzutage auf relativ einfache Weise durchgeführt werden. Für eine exemplarische Studie dieses Typs haben wir die Texte der Wochenzeitschrift „Die ZEIT“ nach den in Abschnitt 4 untersuchten Aktionswörtern (*lol*, *lach*, *freu*, *grins*, *wink*) durchsucht. Wir nutz-

⁴⁶ Artikelseite „Chat“.

ten dafür das ZEIT-Korpus, das als Bestandteil des digitalen lexikalischen Informationssystems DWDS an der Berlin-Brandenburgischen Akademie der Wissenschaften zur Verfügung steht⁴⁷. Das Korpus umfasst alle ZEIT-Ausgaben von 1946 bis 2009 (ca. 460 Millionen Tokens); da erst ab dem Aufkommen des WWW und der damit einhergehenden Popularisierung des Internets mit einem Spracheinfluss zu rechnen ist, wurde nur der Zeitraum 1991–2009 ausgewertet⁴⁸.

Form	Gesamtzahl der Treffer	Pseudotreffer	Metakommunikative Treffer	Echte Verwendungen
lach	133	128	5	0
freu	102	101	1	0
lol	20	4	16	0
wink	17	17	0	0
grins	3	2	1	0

Tabelle 5: Suche nach Aktionswörtern im DWDS-Zeitungskorpus „Die ZEIT“.

An den Ergebnissen in Tabelle 5 überrascht wiederum v.a. die Eindeutigkeit: Zu keinem der Aktionswörter fand sich in den Ausgaben der ZEIT von 1991–2009 auch nur ein einziger echter Beleg. Allerdings zeigt auch diese Untersuchung, dass es unerlässlich ist, die von Korpusrecherchesystemen erzeugten Trefferlisten manuell zu bereinigen, denn selbst wenn die Korpora – wie das DWDS-Korpus – linguistisch aufbereitet sind, ist es auch in ihnen nicht möglich, homographe Formen von Aktionswörtern zuverlässig auszuschließen oder gar metakommunikative Treffer wie das Beispiel 5 automatisch zu erkennen.

- (5) Die Rede ist von DLS mit und ohne Fastpath, von Ping-Raten und Verzögerung in der Datenübermittlung. Jeder zweite Satz wird mit » Lol « oder » Rofl « beendet. Lol?⁴⁹

In allen Trefferlisten fanden sich dieselben Typen von Pseudotreffern wie in der Wikipedia-Auswertung. Die relativ hohe Anzahl von metakommunikativen Treffern zu „lol“ belegt, dass dieses Akronym auch in der journalistischen Diskussion um die Sprachlichkeit in der internetbasierten Kommunikation als typisches Stilelement hervorgehoben wird.

6. Fazit und Ausblick

Die korpusbasierten Untersuchungen, die in diesem Artikel diskutiert wurden, untermauern allesamt die in der Linguistik vertretene Hypothese des sprachlichen Ausbaus und die damit verbundene Auffassung, dass Schreiben und Schriftlichkeit durch die neuen interaktionsorientierten Schreibtechnologien nicht beschädigt werden, sondern sich im Hinblick auf neue Anforderungen erweitern und neu ausdifferenzieren. Bislang gibt es keine empirischen Indizien dafür, dass die Versprachlichungsmuster des interaktionsorientierten Schreibens den Sprach-

⁴⁷ Das Korpus ist verfügbar unter <http://www.dwds.de>. Wir suchten nach den Formen „@lol“, „@lach“, „@freu“, „@grins“ – eine Suche nach dem Verbstamm ohne Operator würde zusätzlich alle flektierten Formen des Verbs ausgeben. Für die Suche nach dem Aktionswort „wink“ haben wir die komplexere Anfrage „@wink &&! =NN with @wink“ genutzt, um Pseudotreffer zum Nomen „der Wink“ auszuschließen.

⁴⁸ Hierfür setzten wir den DWDS-Suchfilter auf den Zeitraum 1990-12-31 bis 2009-12-31.

⁴⁹ Aus: Netz-Opa mit Ende 30. DIE ZEIT, 01.05.2008, Nr. 19. <http://www.zeit.de/2008/19/I-Treffen-Second-Life>.

stil beim textorientierten Schreiben beeinflussen. In dem in Abschnitt 5 untersuchten Zeitungskorpus fand sich in den 1990-2009 erschienenen Ausgaben der Wochenzeitschrift „Die ZEIT“ kein einziger Beleg für die untersuchten einschlägigen Stilmarker. Die in Abschnitt 4 diskutierte Vergleichsstudie, in der Stilmarker in Artikel- und Diskussionsseiten der deutschen Wikipedia ausgewertet wurden, dokumentiert, dass die Akteure in sozialen Netzwerken offensichtlich sehr gut in der Lage sind, zwischen text- und interaktionsorientiertem Schreiben zu unterscheiden. Dass dieses Differenzierungsvermögen bereits bei jugendlichen Schreibern im Grundsatz vorhanden ist, hat die umfangreiche empirische Studie zu schulischen und privaten Schreibprodukten Schweizer Jugendlicher ergeben (Dürscheid et al. 2010). Vieles spricht also dafür, dass kompetente Schreiber zwischen verschiedenen Schreibhaltungen und den jeweiligen Stilen wechseln können. Für das interaktionsorientierte Schreiben werden allerdings andere Kompetenzen benötigt als für das Verfassen elaborierter Schrifttexte. In der Sprach- und Schreibdidaktik wird es deshalb künftig auch darum gehen, die Kompetenz zur situativ und kontextuell angemessenen Sprach- und Stilwahl auch im Hinblick auf die neuen interaktionsorientierten Schreibformen zu fördern (vgl. Storrer i.Dr.).

Hierfür ist aber die Einsicht wichtig, dass die charakteristischen Stilmerkmale nicht an das Medium (Internet) oder ein bestimmtes soziales Netzwerk (deutsche Wikipedia) gebunden sind, sondern dass die Schreiber ihren Schreibstil an das jeweilige kommunikative Setting und die dafür üblichen sprachlichen Gepflogenheiten anpassen. Diese Sprach- und Stilvariation wurde in Abschnitt 3 auf der Basis des Dortmunder Chat-Korpus für eine prototypische interaktive Schreibtechnologie, nämlich den Chat, gezeigt. Die in Abschnitt 3.3. diskutierte quantitative Vergleichsstudie liefert empirische Indizien für die Annahme, dass auch beim Chatten mehrere Stilregister zur Verfügung stehen, aus denen die Chatter diejenigen auswählen, die aus ihrer Sicht am besten zum kommunikativen Setting und den darin verankerten Rollenkonstellationen passen. Da oft mehrere Alternativen für dieselbe sprachliche Handlung zur Verfügung stehen, können die Chatter einen individuellen „Chatstil“ entwickeln, der neben dem Nickname zur Ausbildung einer vornehmlich schriftsprachlich konstituierten „Chatter-Persönlichkeit“ beiträgt. In Abschnitt 3.4 wurden Ergebnisse aus einer von Kristin Luckhardt durchgeführten Untersuchung zum Sprachstil in einer Chat-Community vorgestellt (Luckhardt 2009), an denen solche individuellen Präferenzen für bestimmte stilistische Handlungsmuster auch mit quantitativen Analysen herausgearbeitet wurden.

Linguisten werden von den Ergebnissen der korpuslinguistischen Studien nicht überrascht sein; bemerkenswert ist allenfalls die Eindeutigkeit der Resultate, die bei der Analyse des umfangreichen Zeitungskorpus und beim Vergleich der Vorkommensfrequenzen von einschlägigen Stilmarkern in den beiden Seitentypen der Wikipedia erzielt wurden. Die Studien in Abschnitt 4 und 5 nutzten ein relativ kleines Set von einschlägigen Stilmarkern, die wegen ihrer speziellen Form auch mit vertretbarem Zeitaufwand in sehr großen Korpora untersucht werden konnten. Bereits bei diesen recht gut formal identifizierbaren Einheiten wurde aber in beiden Studien sehr deutlich, dass die Trefferlisten unbedingt manuell gesichtet und um Pseudotreffer bereinigt werden müssen, um aussagekräftige Vorkommenszahlen zu erhalten. Um die sprachlichen Unterschiede zwischen interaktionsorientiertem und textorientiertem Schreibduktus auf allen relevanten Ebenen – Orthographie, Morphologie, Syntax, Lexik (Stilregister) – empirisch herauszuarbeiten, würde man Korpora benötigen, die linguistisch wesentlich feiner annotiert sind. Die in Abschnitt 3 diskutierten Studien am Dortmunder Chat-Korpus zeigen die erweiterten Auswertungsmöglichkeiten, die sich durch strukturelle und linguistische Annotationen für quantitative Analysen ergeben.

Bereits die sehr simple Vergleichsstudie in Abschnitt 4 macht deutlich, dass der Vergleich von Artikel- und Diskussionsseiten sehr gut geeignet ist, um die Variation zwischen beiden Schreibhaltungen in derselben Community an großen Datenmengen zu untersuchen. Im Hinblick auf die in Abschnitt 2 erläuterte Differenzierung von Nähe (konzeptionelle Mündlichkeit) und Distanz (konzeptionelle Schriftlichkeit) lässt sich an vielen Merkmalen nachweisen, dass die Diskussionsbeiträge Charakteristika konzeptioneller Mündlichkeit aufweisen, während die Artikelseiten im Duktus der konzeptionellen Schriftlichkeit verfasst sind. Um solche Unterschiede auch quantitativ untersuchen zu können, wäre allerdings eine linguistisch aufbereitete, zumindest lemmatisierte und wortartenannotierte Version der beiden Seitentypen erforderlich. Auch wäre es sinnvoll, auf die Dokumentstrukturen der Artikelseiten bzw. die Threadstrukturen der Diskussionsseiten (inkl. der Metadaten zu Autoren und Timestamps) gezielt zugreifen zu können. Da die Wikipedia in sehr vielen Kontexten eine sehr interessante Ressource ist, kann man davon ausgehen, dass derartige aufbereitete Versionen in absehbarer Zukunft für die linguistische Forschung zur Verfügung stehen⁵⁰.

Um die Sprachentwicklung und Sprachvariation in den verschiedenen digitalen Genres systematisch untersuchen zu können, wären linguistisch annotierte Korpora wünschenswert, die über eine Zeitspanne hinweg verschiedene Formen internetbasierter Kommunikation in ausgewogenem Verhältnis berücksichtigen. Bislang gibt es lediglich kleinere Korpora, die im Rahmen von Einzelprojekten aufgebaut wurden; auch fehlt es an standardisierten Verfahren zur linguistischen Annotation derartiger Korpora (Beißwenger und Storrer 2008). Der Aufbau von ausgewogenen, linguistisch aufbereiteten Korpora zur internetbasierten Kommunikation und ihre Bereitstellung über Online-Schnittstellen von Korpusinfrastrukturprojekten könnten erheblich dazu beitragen, die Forschung zu den neuen Entwicklungen im Umgang mit Schriftlichkeit auf ein solides und überprüfbares empirisches Fundament zu stellen.

7. Literatur

- Ágel, Vilmos, und Mathilde Hennig (Hrsg.). 2006. *Grammatik aus Nähe und Distanz. Theorie und Praxis am Beispiel von Nähetexten 1650-2000*. Tübingen: Niemeyer.
- Ágel, Vilmos, und Mathilde Hennig (Hrsg.). 2010. *Nähe und Distanz im Kontext variationslinguistischer Forschung*. Berlin und New York: de Gruyter.
- Androutsopoulos, Jannis K., und Evelyn Ziegler. 2003. Sprachvariation und Internet: Regionalismen in einer Chat-Gemeinschaft. In *Standardfragen: soziolinguistische Perspektiven auf Sprachgeschichte, Sprachkontakt und Sprachvariation*, Hrsg. Jannis K. Androutsopoulos und Evelyn Ziegler, 251-279. Frankfurt a. M.: Lang.
- Beißwenger, Michael. 2000. *Kommunikation in virtuellen Welten: Sprache, Text und Wirklichkeit*. Stuttgart: ibidem.
- Beißwenger, Michael (Hrsg.). 2001a. *Chat-Kommunikation. Sprache, Interaktion, Sozialität & Identität in synchroner computervermittelter Kommunikation. Perspektiven auf ein interdisziplinäres Forschungsfeld*. Stuttgart: ibidem.
- Beißwenger, Michael. 2007. *Sprachhandlungskoordination in der Chat-Kommunikation*. Berlin und New York: de Gruyter.
- Beißwenger, Michael. 2010. *Chattern unter die Finger geschaut: Formulieren und Revidieren bei der schriftlichen Verbalisierung in synchroner internetbasierter Kommunikation*. In *Nähe und Distanz im Kontext variationslinguistischer Forschung*, Hrsg. Vilmos Ágel und Mathilde Hennig, 247-294. Berlin und New York: de Gruyter.

⁵⁰ Vgl. für die englische Wikipedia Ferschke/Gurevych/Chebotar 2012.

- Beißwenger, Michael. in diesem Band: Raumorientierung in der Netzkommunikation. Korpusgestützte Untersuchungen zur lokalen Deixis in Chats.
- Beißwenger, Michael, und Angelika Storrer. 2005. Chat-Szenarien für Beruf, Bildung und Medien. In *Chat-Kommunikation in Beruf, Bildung und Medien: Konzepte – Werkzeuge – Anwendungsfelder*, Hrsg. Michael Beißwenger und Angelika Storrer, 9-25. Stuttgart: ibidem.
- Beißwenger, Michael, und Angelika Storrer. 2008. Corpora of Computer-Mediated Communication. In *Corpus Linguistics. An International Handbook*, Band 1, Hrsg. Anke Lüdeling und Merja Kytö, 292-308. Berlin und New York: de Gruyter.
- Brinker, Klaus, Gerd Antos, Wolfgang Heinemann und Sven F. Sager (Hrsg.). 2000. *Text- und Gesprächslinguistik. / Linguistics of Text and Conversation. Ein internationales Handbuch zeitgenössischer Forschung*. 1. Halbband: Textlinguistik. Berlin und New York: de Gruyter.
- Brinker, Klaus, Gerd Antos, Wolfgang Heinemann, und Sven F. Sager (Hrsg.). 2001. *Text- und Gesprächslinguistik. / Linguistics of Text and Conversation. Ein internationales Handbuch zeitgenössischer Forschung*. 2. Halbband: Gesprächslinguistik. Berlin und New York: de Gruyter.
- Brommer, Sarah. 2007. „Ein unglaubliches Schriftbild, von Rechtschreibung oder Interpunktion ganz zu schweigen“ – Die Schreibkompetenz der Jugendlichen im öffentlichen Diskurs. *Zeitschrift für germanistische Linguistik* 35 (3), 315-345.
- Diekmannshenke, Hajo. 2005. Politische Kommunikation in Zeiten des Internet. Kommunikationswandel am Beispiel moderierter und unmoderierter Politik-Chats. In *Chat-Kommunikation in Beruf, Bildung und Medien: Konzepte – Werkzeuge – Anwendungsfelder*, Hrsg. Michael Beißwenger und Angelika Storrer, 119-144. Stuttgart: ibidem.
- Diewald, Nils, Frank-Job, Barbara. in diesem Band. Kollaborative Aushandlung von Kategorien am Beispiel des GutenPlag Wikis.
- Dorta, Gabriel. 2005. *Soziale Welten in der Chat-Kommunikation. Untersuchungen zur Identitäts- und Beziehungsdimension in Web-Chats*. Bremen: Hempen.
- Dürscheid, Christa, und Sarah Brommer. 2009. Getippte Dialoge in neuen Medien. Sprachkritische Aspekte und linguistische Analysen. *Linguistik online* 37 (1).
- Dürscheid, Christa, Franc Wagner, und Sarah Brommer. 2010. *Wie Jugendliche schreiben. Schreibkompetenz und neue Medien*. Berlin und New York: de Gruyter.
- Ehlich, Konrad. 1994. Funktion und Struktur schriftlicher Kommunikation. In *Schrift und Schriftlichkeit. Ein interdisziplinäres Handbuch internationaler Forschung*, Band 1, Hrsg. Hartmut Günther und Otto Ludwig, 18-41. Berlin und New York: de Gruyter.
- Feldweg, Helmut, Ralf Kibiger, und Christine Thielen. 1995. Zum Sprachgebrauch in deutschen Newsgruppen. *Osnabrücker Beiträge zur Sprachtheorie* 50, 143-154.
- Frank-Job, Barbara (2009): Formen und Folgen des Ausbaus französischer Nähesprache in computervermittelter Kommunikation. In: Pfänder, Stefan / Kailuweit, Rolf / Cousin, Vanessa (Hgg.), *FrankoMedia: Aufriss einer französischen Sprach- und Medienwissenschaft*. Berlin: Berliner Wissenschafts-Verlag, 71-88.
- Frank-Job, Barbara (2010): Sprachwandel und Medienwandel. In: Mehler, Alexander; Sutter, Tilmann (Hg.), *Medienwandel als Wandel von Interaktionsformen — von frühen Medienkulturen zum Web 2.0*. Wiesbaden: Verlag für Sozialwissenschaften, 27-46.
- Ferschke, Oliver; Gurevych, Iryna; Chebotar, Yevgen (2012): Behind the Article: Recognizing Dialog Acts in Wikipedia Talk Pages. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012).

- Günthner, Susanne, und Gurly Schmidt. 2002. Stilistische Verfahren in der Welt der Chat-groups. In *Soziale Welten und kommunikative Stile*, Hrsg. Inken Keim und Wilfried Schütte, 315-337. Tübingen: Narr.
- Haase, Martin, Michael Huber, Alexander Krumeich, und Georg Rehm. 1997. Internetkommunikation und Sprachwandel. In *Sprachwandel durch Computer*, Hrsg. Rüdiger Weingarten, 51-85. Opladen: Westdeutscher Verlag.
- Herring, Susan. 2010. Computer-Mediated Conversation: Introduction and Overview. *Language@Internet* 7. Online: <http://www.languageatinternet.de/articles/2010/2801>
- Hoffmann, Ludger. 2004. Chat und Thema. *Osnabrücker Beiträge zur Sprachtheorie* 68, 103-122.
- Kilian, Jörg. 2001. T@stentöne. Geschriebene Umgangssprache in computervermittelter Kommunikation. In *Chat-Kommunikation. Sprache, Interaktion, Sozialität & Identität in synchroner computervermittelter Kommunikation. Perspektiven auf ein interdisziplinäres Forschungsfeld*, Hrsg. Michael Beißwenger, 55-78. Stuttgart: ibidem.
- Koch, Peter, und Wulf Oesterreicher. 1994. Schriftlichkeit und Sprache. In *Schrift und Schriftlichkeit. Ein interdisziplinäres Handbuch internationaler Forschung*, Band 1, Hrsg. Hartmut Günther und Otto Ludwig, 587-604. Berlin und New York: de Gruyter.
- Koch, Peter, und Wulf Oesterreicher. 2007. Schriftlichkeit und kommunikative Distanz. *Zeitschrift für germanistische Linguistik* 35 (3), 346-375.
- Luckhardt, Kristin. 2009. *Stilanalysen zur Chat-Kommunikation. Eine korpusgestützte Untersuchung am Beispiel eines medialen Chats*. Dissertation, TU Dortmund. <http://hdl.handle.net/2003/26055>.
- Lemnitzer, Lothar, und Karin Naumann. 2001. „Auf Wiederlesen!“ – das schriftlich verfasste Unterrichtsgespräch in der computervermittelten Kommunikation. In *Chat-Kommunikation. Sprache, Interaktion, Sozialität & Identität in synchroner computervermittelter Kommunikation. Perspektiven auf ein interdisziplinäres Forschungsfeld*, Hrsg. Michael Beißwenger, 469-491. Stuttgart: ibidem.
- Mehler, Alexander, Serge Sharoff, und Marina Santini (Hrsg.). 2010. *Genres on the Web: Computational Models and Empirical Studies*. Dordrecht: Springer.
- Polenz, Peter v. 1999. *Deutsche Sprachgeschichte vom Spätmittelalter bis zur Gegenwart*. Band 3. 19. und 20. Jahrhundert. Berlin und New York: de Gruyter.
- Raible, Wolfgang 1999. Kognitive Aspekte des Schreibens. Heidelberg, C.Winter
- Runkehl, Jens, Peter Schlobinski und Torsten Siever. 1998. *Sprache und Kommunikation im Internet. Überblick und Analysen*. Opladen: Westdeutscher Verlag.
- Sandig, Barbara. 1986. *Stilistik der deutschen Sprache*. Berlin und New York: de Gruyter.
- Sandig, Barbara. 2006. *Textstilistik des Deutschen*. Berlin und New York: de Gruyter.
- Schlobinski, Peter (Hrsg.). 2006. *Von *hdl* bis *cul8r**. *Sprache und Kommunikation in neuen Medien*. Mannheim: Dudenverlag.
- Schmitz, Ulrich. 2004. *Sprache in modernen Medien. Einführung in Tatsachen und Theorien, Themen und Thesen*. Berlin: Erich Schmidt.
- Sieber, Peter. 1998. *Parlando in Texten. Zur Veränderung kommunikativer Grundmuster in der Schriftlichkeit*. Tübingen: Niemeyer.
- Stegbauer, Christian. 2009. *Wikipedia. Das Rätsel der Kooperation*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Storrer, Angelika. 2001. Getippte Gespräche oder dialogische Texte? Zur kommunikationstheoretischen Einordnung der Chat-Kommunikation. In *Sprache im Alltag. Beiträge zu*

- neuen Perspektiven in der Linguistik*, Hrsg. Andrea Lehr, Matthias Kammerer, Klaus-Peter Konerding et al., 439-465. Berlin und New York: de Gruyter.
- Storrer, Angelika. 2007. Chat-Kommunikation in Beruf und Weiterbildung. *Der Deutschunterricht* 2007 (1), 49-61.
- Storrer, Angelika. i.Dr.. Neue Text- und Schreibformen im Internet: Das Beispiel Wikipedia. Erscheint in: Köster, Juliane, Feilke, Helmuth (Hg.): *Textkompetenzen in der Sekundarstufe II*. Freiburg: Fillibach.
- Thaler, Verena. 2003. *Chat-Kommunikation im Spannungsfeld zwischen Oralität und Literalität*. Berlin: VWF-Verlag.
- Zesch, Torsten, Christoph Müller, und Iryna Gurevych. 2008. Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary. *Proceedings of the Conference on Language Resources and Evaluation (LREC), Marrakesch, Marokko*.
- Zifonun, Gisela, Ludger Hoffmann, Bruno Strecker u.a. (Hrsg.). 1997. *Grammatik der deutschen Sprache*. 3 Bde. Berlin und New York: de Gruyter.