



**tu** technische universität  
dortmund

# Das Dortmunder Chat-Korpus



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

**Michael Beißwenger - Angelika Storrer**

**tu** Dortmund

Institut für deutsche Sprache und Literatur

[michael.beisswenger@udo.edu](mailto:michael.beisswenger@udo.edu)

[angelika.storrer@udo.edu](mailto:angelika.storrer@udo.edu)



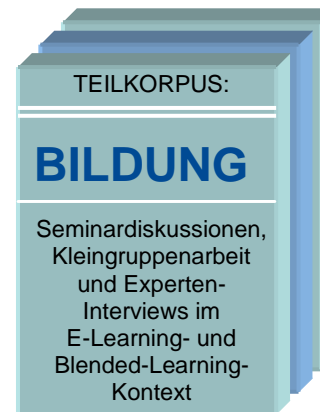
Projekt am Institut für deutsche Sprache und Literatur,  
Lehrstuhl für Linguistik der deutschen Sprache und Sprachdidaktik (seit 2002)

BESTAND: 480 Mitschnitte mit insgesamt  
140.558 Chat-Beiträgen bzw. 1,07 Mio. Tokens

90 Mitschnitte  
88.262 Chat-Beiträge  
517.828 Tokens



49 Mitschnitte  
13.240 Chat-Beiträge  
92.528 Tokens

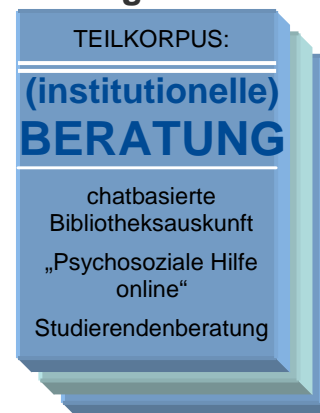


Chat-Daten aus Anwendungen in unter-  
schiedlichen sozialen Handlungsbereichen

99 Mitschnitte  
17.716 Chat-Beiträge  
237.767 Tokens



242 Mitschnitte  
21.340 Chat-Beiträge  
219.345 Tokens



Projekt am Institut für deutsche Sprache und Literatur,  
Lehrstuhl für Linguistik der deutschen Sprache und Sprachdidaktik (seit 2002)

BESTAND: 480 Mitschnitte mit insgesamt  
140.558 Chat-Beiträgen bzw. 1,07 Mio. Tokens

Gesamtbestand **XML-**  
**annotiert** (Typen von Chat-  
Beiträgen; chat-typische  
Stilelemente; Metadaten)

80% davon frei zugänglich



Freigegebener  
KorpUSAusschnitt  
(385 Dokumente) zusätzlich  
als **HTML-Version**



## Chat im Freizeitbereich

SPOOKY Irgendwie ist jetzt an mir was vorbeigeschossen

Findalf Hausdrache, nö, und ja, er ist scheiß langsam!

Arktikus GF: \*ggg\*...hmm..der aiuch...auff jden Fall zu KArneval \*s\*

desertstorm betritt den Raum.

*ruebennase langweilt sich immer noch....*

GF: 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

## Chat im Handlungsbereich „Beratung“

- BENUTZER* Können Sie mir sagen, ob das Buch Fn 25665 ausgeliehen ist?
- 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* Benutzer - 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.
- 11:56:00 *BENUTZER* Vielen Dank!
- 11:56:09 *AUSKUNFT* Gern geschehen und schönen Tag noch.
- 11:56:16 \*\*\* *BENUTZER* hat den chat verlassen. \*\*\*

Search The Corpus!

www.chatkorpus.uni-dortmund.de  
STACCA<sup>Do</sup>  
chat korpus dortmund 1.0

### Grundeinstellungen

Wählen Sie die zu durchsuchende Datei bzw. den zu durchsuchenden Ordner:   
Ausgewählter Pfad: *Chat Korpus/*

Wählen Sie den Pfad zur Output-Datei:  
Ordner:       Dateiname:

Wählen Sie den Message-Kontext:

### Suchanfrage

Spezifizieren Sie Ihre Suchanfrage:

Suchstring:

Die Message soll außerdem folgende Elemente enthalten:

Nickname       allgemein       nur Benutzer

Emoticon

Asterisk Expression

Adressierung       allgemein       an Benutzer

Zeige nur Messages ...

vom Typ "system"

vom Typ "utterance"

vom Typ "action"

vom Benutzer

### Suchoptionen

Mitschnitt	TNOM	NOM Human	TNOT	NOT Human	TNOC	NOC Human	Ø-Länge Message	# Em.	
<b>Datei:</b> 2221007_unicum_21-02-2003_(2).xml <b>(Teilkorpus FREIZEIT: Webchat bei unicum.de)</b>	531	452	2251	1831	12852	10295	4,0508847	31	...
<b>Datei:</b> 1102020_seminar_classroom_04-12-2002.xml <b>(Teilkorpus BILDUNG: chatbasiertes Seminar)</b>	349	273	2315	1863	15273	12218	6,824176	4	
<b>Datei:</b> 1301005_politik-digital_Guenther_Beckstein_19-05-2004.xml <b>(Teilkorpus MEDIEN: Politiker-Chat mit Günther Beckstein bei politik-digital.de)</b>	79	78	1923	1881	13548	13275	24,115385	1	
<b>Datei:</b> 1204014_14_2005-01-12_aver.xml <b>(Teilkorpus BERATUNG: 1:1-Beratung zu Online-Auktionen)</b>	172	163	1923	1863	11476	11177	11,429448	10	

... zum Beispiel als Grundlage für Untersuchungen zur Frequenz bestimmter Stilelemente in unterschiedlichen Handlungsbereichen (z.B. Storrer 2007)

Nickname	est. Gender	NOM	NOT	NOC	Ø-Länge Message	% Messages	% Tokens	% Characters	Länge/Ø-Chatter
McMike	male	111	573	4033	5,1621623	1,85	2,222136	2,6964686	120,11547
marc30	male	114	483	2774	4,236842	1,9	1,8731095	1,8546996	98,58471
zora	female	121	768	4154	6,3471074	2,0166667	2,9783602	2,777369	147,68729
ineli26	female	128	671	4144	5,2421875	2,1333332	2,6021872	2,7706833	121,97753
Dill	male	145	425	2305	2,9310346	2,4166665	1,6481811	1,5411257	68,20061
mr.floppy	male	145	420	2305	2,8965516	2,4166665	1,6287909	1,5411257	67,39824
Rosenstaub1979	female	153	419	2336	2,738562	2,55	1,6249127	1,5618522	63,72207
Weswolf	male	159	655	3720	4,119497	2,6499999	2,540138	2,4871962	95,85427
Tigerelse	female	168	619	3261	3,6845238	2,8000002	2,4005272	2,1803083	85,73312
einer	male	224	1227	6962	5,477679	3,7333333	4,758396	4,6548014	127,45705
Lantonie	female	300	1120	6353	3,7333333	5	4,3434424	4,247623	86,86884
quaki	female	340	1065	6161	<b>3,1323528</b>	5,666667	4,130148	4,1192517	72,884964
TomcatMJ	male	384	2604	17599	<b>6,78125</b>	6,4	10,098503	11,766711	157,78911
stoeps	female	483	1837	10069	3,8033125	8,05	7,124021	6,732145	88,497154

... zum Beispiel als Grundlage für Untersuchungen zum individuellen Chat-Verhalten bzw. zu individuellen Chat-Stilen (z.B. Luckhardt 2009)





## 18 Chats im Handlungsbereich

„Beratung“ mit insgesamt 32  
ProbandInnen von jeweils 32-44 Minuten  
Dauer (Konstellation 1:1, 1:2, 1:3 und 1:4)

## Erhobene Daten:

- 11 Std 26 Min Mitschnitte
- 25 Std 13 Min „screen movies“  
(Bildschirmvideos)
- 28 Std 43 Min Videobeobachtung

## Korpus zur kommunikativen Teilhabe beim Chatten (Beißwenger 2007, 2009)

14:41:11

Talk to: all  
 show for who  
 show server messages  
 show timestamps

ich hätte eine Frage zum Rückgaberecht bei mangelhaft

server: briha entered the room ebay6 at 14:34:48  
 server: nabro entered the room ebay6 at 14:40:10  
 server: bsommer entered the room ebay6 at 14:40:22  
 bsommer: Hallo und herzlich willkommen zur eBay-Onlineberatung! Ich  
 bin Bianca Sommer und werde versuchen, Ihnen einige Fragen zum  
 Thema eBay zu beantworten, auf die Sie schon immer eine Antwort  
 wissen wollten. ;-)  
 briha: Hallo Frau Sommer, hallo nabro  
 bsommer: Gibt es spezielle Fragen, die Ihnen schon unter den Nägeln  
 brennen?

Four video thumbnails showing participants: a man in a yellow sweater, a woman with blonde hair, and two other views of the same participants.

### Desiderate/Perspektiven in Bezug auf die Erhebung und Aufbereitung von multimodalen Daten:

- a) Technische Alignierung von „Produkt-“ und „Prozessdaten“ (Mitschnitt – „screen movie“ – Videoaufzeichnung);
- b) Einbeziehung weiterer Erhebungsmethoden (z.B. [Eye Tracking](#)).

### Ausbau des Korpus:

Erweiterung des Korpus um Daten auch zu anderen (auch audio-/video-basierten) Genres im Bereich der synchronen *Computer-Mediated Communication*:

- **Instant Messaging** (ICQ, AOL IM etc.)
- **Voice-over-IP-Anwendungen** (Skype, Teamspeak etc.)
- **Videoconferencing**

sowie zu Applikationen, die textbasierte, audio- und video-gestützte Kommunikationsfunktionen integrieren und kombiniert nutzbar machen.

### Desiderate in Bezug auf die Aufbereitung von Mitschnittsdaten für linguistische Recherchezwecke:

- a) Vorverarbeitungs-Tools und Suchtechniken, die mit nicht normgemäßen Schreibungen (liberaler Umgang mit der Orthographie; Verfremdungs-Schreibungen), Tippfehlern, Kurzformen und netzspezifischen Sonderelementen umgehen können;
- b) Methoden/Werkzeuge, die eine semi-automatische Rekonstruktion von Diskursstrukturen („threads“) unterstützen;
- c) Klärung von Fragen des Urheberrechts in Bezug auf Chat-Daten;
- d) Methoden/Werkzeuge, die eine semi-automatische Identifizierung und Maskierung/Demaskierung personenbezogener Daten in Chat-Mitschnitten erlauben (Ziel: automatische, reversible Anonymisierung von Nicknames, erwähnten Orts- und Personennamen)



technische universität  
dortmund

# Das Dortmunder Chat-Korpus

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

- M. Beißwenger & A. Storrer (2008): [Corpora of Computer-Mediated Communication](#). In: Anke Lüdeling & Merja Kytö (Eds): *Corpus Linguistics. An International Handbook*. Volume 1. Berlin. New York: de Gruyter (Handbücher zur Sprache und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science 29.1), 292-308.
- M. Beißwenger (2007): [Sprachhandlungskoordination in der Chat-Kommunikation](#). Berlin. New York: de Gruyter (Linguistik – Impulse & Tendenzen 26). [Zugl. Diss., TU Dortmund]
- M. Beißwenger (2009): [Chattern unter die Finger geschaut: Formulieren und Revidieren bei der schriftlichen Verbalisierung in synchroner internetbasierter Kommunikation](#). In: Vilmos Ágel & Mathilde Hennig (Hrsg.): *Nähe und Distanz im Kontext variationslinguistischer Forschung*. Berlin. New York: de Gruyter (Linguistik – Impulse & Tendenzen). [erscheint]
- K. Luckhardt (2009): [Stilanalysen zur Chat-Kommunikation. Eine korpusgestützte Untersuchung am Beispiel eines medialen Chats](#). Diss., TU Dortmund. Digitale Ressource: <http://hdl.handle.net/2003/26055>
- A. Storrer (2009): [Rhetorisch-stilistische Eigenschaften der Sprache des Internets](#). In: Ulla Fix, Andreas Gardt & Joachim Knappe (Hgg.): *Rhetorik und Stilistik – Rhetorics and Stilistics*. Berlin. New York: de Gruyter (Handbücher zur Sprache und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science). [erscheint]
- A. Storrer (2007): [Chat-Kommunikation in Beruf und Weiterbildung](#). In: *Der Deutschunterricht* 1/2007, 49-61.