

480 Chat-Mitschnitte in unterschiedlichen Kompilationen:

Korpus	Dokumente	Chat-Beiträge	Tokens	Chat-Beiträge		Tokens	
				(bereinigt von Systemmeldungen)			
Basiskorpus	480	140.558	1.067.468				
Releasekorpus	385	59.876	551.762				
BalaCK 1a	149	10.045	71.661			4 x 15.000	(60.000)
BalaCK 1b	162	9.215	75.124	4 x 2.000	(8.000)		
BalaCK 2a	201	26.669	200.828			4 x 45.000	(180.000)
BalaCK 2b	209	27.066	237.686	4 x 6.000	(24.000)		



Vergleichskompilationen:
BalaCK
 („Balanced“ Chat-Korpora, ausgewogen nach Handlungsbereichen)

verfügbar als (a) **downloadbare XML-Version** mit Abfragewerkzeug **STACADO** und als (b) **online nutzbare HTML-Volltextsammlung** (z.B. als Ressource für Lehrerinnen und Lehrer)



spezialisiertes Abfragewerkzeug

Seiltänzer Zora hat Geburtstag?
quaki wer is glimsey??
Lantonie Ja, und wie!
stoeps ja seil
Seiltänzer *allesverpenn*
basti26 was hast du denn für ein auto?
Seiltänzer Herzlichsten Glückwunsch, zora :) *handschüttel*
zora hat ein auto mit neuen Querlenkern :/

```
<message id="1770" type="utterance" creator="Seiltänzer" color="#6600DD">
  <messageHead>
    < nickname>Seiltänzer</ nickname>
  </messageHead>
  <messageBody>
    Herzlichsten Glückwunsch,
    <address addresssee="zora">zora</address>
    <emoficon>:o:</emoficon>
    <asteriskExpression>*handschüttel*
  </messageBody>
</message>

<message id="1771" type="action" creator="zora" color="#FF00A0">
  <messageBody>
    < nickname>zora</ nickname> hat ein auto mit neuen Querlenkern <emoficon>:/</ emoficon>
  </messageBody>
</message>
```

- Annotation:**
- Der einzelne Chat-Beitrag (*message*) als Grundeinheit
 - Typen von *messages*:
 - „utterance“ (Standard-Beitrag)
 - „action“ (Beitrag mit Zuschreibungscharakter)
 - „system“ (Systemmeldung)
 - Kategorien für chat-typische Stilelemente:
 - Nicknames
 - Ausdrücke in Asterisken
 - Adressierungen
 - Metadaten zum vermeintlichen Geschlecht und zum Beitragsverhalten der dokumentierten Chatter (Anzahl und Länge von Beiträgen im Verhältnis zum Durchschnitt)

stringbasierte Abfrage (mit regulären Ausdrücken)
Suche nach Kategorien
Erzeugung statistischer Übersichten

ANWENDUNGSBEISPIEL

Forschungsfrage: Kommen die häufig als „typisch chatsprachlich“ beschriebenen Elemente *Emoticons* und *Asterisk-Ausdrücke* in allen Anwendungskontexten gleich häufig vor?

	Freizeit	Bildung	Medien	Beratung	insgesamt
Emoticons	546 46,7 %	316 27,0 %	123 10,5 %	186 15,8 %	1170
Asterisk expressions	1056 87,1 %	41 3,4 %	105 8,7 %	10 0,8 %	1212

Chatter-Profil / Teilkorpus „Freizeit“ aus BalaCK 2b (Ausschnitt):

Nickname	est. Gender	NOM	NOT	NOC	Ø-Länge Message	% Messages	% Tokens	% Characters	LängeØ Chatter
McMike	male	111	573	4033	5.1621623	1,85	2.222136	2.8964686	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
inell26	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.flippy	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
Wesewolf	male	159	655	3720	4.119497	2.6499999	2.540138	2.4871962	95.85427
Tigerreise	female	168	619	3261	3.6845238	2.8000002	2.4005272	2.1803083	85.73312
einer	male	224	1227	6962	5.477679	3.733333	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	3.1323528	5.6666667	4.130148	4.1192517	72.884964
TomcatMJ	male	384	2604	17569	6.78125	6,4	10.098503	11.766711	157.78911
stoeps	female	483	1837	10069	3.8033125	8,05	7.124021	6.732145	88.497154

- ANWENDUNGSBEISPIEL**
- Forschungsfragen:**
- Wie lang sind Chat-Beiträge (in Plauder-Chats) im Durchschnitt?
 - Fassen sich (vermeintliche) Männer im Chat kürzer als Frauen?
 - Wie häufig verwenden „Vielchatter“ (in Plauder-Chats) Emoticons / Asterisk-Ausdrücke / Adressierungen und „action messages“?

Logfile-Profil für vier Mitschnitte aus den Bereichen „Freizeit“, „Bildung“, „Medien“ und „Beratung“ (vergleichbare Anzahl an „human tokens“):

Mitschnitt	TNOM	NOM Human	TNOT	NOT Human	TNOC	NOC Human	Ø-Länge Message	# Em.	Messages/Em.	# AstEx.	Messages/AstEx.
Datei: 2221007_unicum_21-02-2003_(2).xml (Teilkorpus FREIZEIT: Webchat bei unicum.de)	531	452	2251	1831	12852	10295	4,0508847	31	17,129632	69	7,695652
Datei: 1102020_seminar_classroom_04-12-2002.xml (Teilkorpus BILDUNG: chatbasiertes Seminar)	349	273	2315	1863	15273	12218	6.824176	4	87,25	1	349,0
Datei: 1301005_politik-digital_Guenter_Beckstein_19-05-2004.xml (Teilkorpus MEDIEN: Politiker-Chat mit Günther Beckstein bei politik-digital.de)	79	78	1923	1881	13548	13275	24.115385	1	79,0	0	0
Datei: 1204014_14_2005-01-12_aver.xml (Teilkorpus BERATUNG: 1:1-Beratung zu Online-Auktionen)	172	163	1923	1863	11476	11177	11.429448	10	17,2	0	0

Nickname	est. Gender	NOM	# Em.	msg/Em.	# Ast. Expr.	msg/Ast. Expr.	Address.	msg/Address.	action msg	msg/action msg
stoeps	female	483	40	12,075	164	2.945121951	162	2.981481481	52	9.288461538
TomcatMJ	male	384	104	3.69230769	115	3.339130435	140	2.742857143	11	34.90909091
quaki	female	340	0	0	36	9.444444444	74	4.594594595	32	10.625
Lantonie	female	300	44	6.81818182	69	4.347826087	57	5.263157895	30	10
einer	male	224	18	12,44444444	49	4.571428571	49	4.571428571	39	5.743589744
Tigerreise	female	168	25	6,72	48	3,5	42	4	2	84
Wesewolf	male	159	23	6.91304348	9	17.66666667	36	4.416666667	22	7.227272727
Rosenstaub1979	female	153	13	11.7692308	24	6.375	25	6.12	4	38.25
mr.flippy	male	145	5	29	16	9.0625	33	4.393939394	6	24.16666667
Dill	male	145	1	145	9	16.11111111	33	4.393939394	14	10.35714286
inell26	female	128	0	0	26	4.923076923	2	64	29	4.413793103
zora	female	121	27	4.48148148	28	4.321428571	32	3.78125	14	8.842857143
marc30	male	114	26	4.38461538	11	10.36363636	56	2.035714286	6	19
McMike	male	111	0	0	18	6.166666667	5	22,2	20	5,55