

Fragen auf Twitter?

Johannes Gontrum & Steve Wendler

31. März 2015

1 Abstract

Diese Arbeit stellt eine korpusbasierte Untersuchung von *Fragen in Twitter-Dialogen* vor. Beschrieben wird, dass auch auf Twitter das Fragezeichen das eindeutigste Erkennungsmerkmal einer Frage ist. Anhand einer Question-Reply-Ratio stellen wir fest, dass der beste Zeitpunkt, auf Twitter eine Frage beantwortet zu bekommen, der Samstagabend ist. Außerdem konnte gezeigt werden, dass die technischen Gegebenheiten auf Twitter zu einer eigenen Form der Generierung von *Echofragen* führt.

2 Einleitung

Twitter ist eine Echtzeitanwendung, die es dem Nutzer erlaubt, von unterschiedlichen Plattformen (Desktop-Computer, Smartphone) maximal 140 Zeichen lange Kurznachrichten zu versenden. Dabei folgt der Dienst gerade nicht dem klassischen Sender-Empfänger-Paradigma der Sprachwissenschaft, sondern eine Nachricht wird für jeden zugänglich der Öffentlichkeit präsentiert. Diese Form der Kommunikation erlaubt dem Sender einer Nachricht ein riesiges Publikum zu erreichen aber auch gänzlich ohne Empfänger zu enden. Interessant für die Sprachwissenschaft ist Twitter aufgrund seines informellen Charakters. Bisher greift die klassische Korpuslinguistik vornehmlich auf Quellen zurück, die sich durch einen gewissenhaften und regelkonformen Umgang mit Sprache auszeichnen (Zeitungen, Parlamentsreden o.ä.). Twitterdaten hingegen zeichnen sich durch einen informellen Schreibstil aus: Benutzer des Dienstes gehen kreativ mit Sprache um, passen sie den technischen Gegebenheiten an und legen oftmals keinen Wert auf formale Regeln. Der soziale Charakter des Mediums führt oftmals dazu, dass bewusst mit Normabweichungen gespielt wird. Gerade hierin besteht für den Sprachwissenschaftler eine große Möglichkeit, synchrone Sprachwissenschaft zu betreiben und Phänomene zu entdecken.

In der vorliegenden Arbeit betrachten wir einen Aspekt von Twitter-Kommunikation: Fragen in Twitter-Dialogen. Twitter erlaubt eine besondere Form der dialogischen Kom-

munikation: durch *Replies* können auf bestimmte Nachrichten Antworten gesendet werden. Durch Analyse von Metadaten können diese Dialoge vollumfänglich rekonstruiert werden.

Uns interessierte, welche Rolle Fragen auf Twitter überhaupt spielen. Wie oft werden Fragen beantwortet? Gibt es Muster der Beantwortung von Fragen? Wenn überhaupt, unterscheiden sich Fragestellungen auf Twitter von normalsprachlichen Fragen.

Um diese Fragen beantworten zu können, war es zunächst notwendig, Dialogstrukturen zu erkennen und durchsuchbar zu machen. In einem weiteren Schritt mussten Fragen erkannt werden.

3 Daten und Vorverarbeitung

3.1 Generierung der Dialogstrukturen

Als Grundlage für die Untersuchungen dient ein Korpus von 20.733.071 deutschen Tweets (exklusive Retweets), die im April 2013 gesammelt wurden. Dieses Korpus wurde mit Hilfe einer deutschen Stoppwortliste durch die Twitter Streaming API generiert und anschließend durch LangID gefiltert, sodass Einträge entfernt wurden, die nicht in deutscher Sprache verfasst wurden.

Um Anfragen an dieses Korpus zu vereinfachen, haben wir das komplette Korpus in eine MySQL Datenbank übertragen und in einem mehrstufigen Prozess für jeden Tweet eine Menge von Attributen erstellt. Dabei wurden einige Attribute direkt aus den JSON-Files der Twitter Streaming API ausgelesen, andere durch neuerliche Iterationen erzeugt.

Twitter Streaming API

- `id_str` – ID des Tweets
- `user_id_str` – ID der Users
- `friend_count` – Anzahl der Personen, denen Tweet-Verfassers folgt
- `followers_count` – Anzahl der Follower des Tweet-Verfassers
- `in_reply_to_status_id_str` – ID des Tweet auf den aktueller Tweet Antwort ist
- `in_reply_to_user_id_str` – ID des Users des Tweet auf den aktueller Tweet Antwort ist
- `user_mentions_count` – Anzahl der User, die im Tweet erwähnt werden
- `user_mentions` – Liste der IDs der User, die im Tweet erwähnt werden
- `created_at` – Zeitpunkt des Erstellens des Tweet

- `source` – von welcher Plattform wurde Tweet gesendet (Web, Smartphone)
- `text` – Textkörper des Tweet

nachbearbeitete Attribute

- `direct_replies_count` – Anzahl der Tweets, die direkt auf den aktuellen Tweet antworten
- `direct_replies_list` – Liste der Tweets, die direkt auf den aktuellen Tweet antworten
- `indirect_replies_count` – Anzahl der Tweets, die im Dialog-Baum unterhalb des aktuellen Tweet hängen
- `indirect_replies_list` – Eine Liste von allen Antworten, die sich auf diesen Tweet bezogen
- `is_base_tweet` – Root-Tweet eines Dialoges
- `is_question` – Tweet ist eine/ keine Frage
- `question_mark` – Anzahl der Fragezeichen eines Tweets
- `is_wh_question` – Tweet beginnt mit einem w-Fragewort
- `valid` – Tweet ist kein Spam.

Auf Grundlage dieser Informationen war es uns möglich, alle Dialoge in jeweils einem *Dialogbaum* abzubilden, um so Fragen und Antworten zu ermitteln. Dieser Dialogbaum wurde aufgebaut, indem alle Tweets auf ihr Attribut `in_reply_to_status_id_str` hin geprüft wurden und so der Tweet, zu dem der aktuelle Tweet eine Antwort darstellt, mit dem aktuellen verbunden werden konnte.

3.2 Basistweets

Ebenso war es uns möglich, Tweets zu markieren, die einen Dialog starten. Diese Basistweets beziehen sich auf keine anderen Tweets, es gibt jedoch Nachrichten, die auf diese Basistweets antworten. Die Identifizierung setzt jedoch voraus, dass NutzerInnen sich bei einer Antwort direkt auf einen Tweet beziehen und nicht manuell eine Nachricht mit einem @-Handle verfassen. So wurde der Tweet

- (1) @DeigningDiamond – wenigstens etwas, das ich machen konnte, ohne dass du es auch nur merken konntest. (ID: 318484529570529281)

von unserem System als Basistweet markiert, betrachtet man aber den Kontext, wird klar, dass er eigentlich Teil einer Diskussion ist. Diese fälschlich markierten Tweets können leider nicht vermieden werden, da es nicht ungewöhnlich ist, einen Dialog mit einem an eine Userin / einen User gerichteten Tweet beginnt:

(2) *Beginn:* @_danjl Dein Bild ist richtig scheiße (ID: 318482949844660224)

Antwort: @chrisgoescross Welches soll ich denn sonst nehmen? (ID: 318483248978219008)

3.3 Direkte und indirekte Antworten

Eine direkte Antwort ist ein Tweet, der sich direkt auf den Tweet bezieht, auf welchen er mit seinem Attribut `in_reply_to_status_id_str` zeigt, während indirekte Antworten transitiv auch Tweets bezeichnen, die auf eine Nachricht antworten, die wiederum eine Antwort auf den Basistweet ist. Zur schnelleren Analyse haben wir ebenfalls die Anzahl der direkten und indirekten Antworten in die Datenbank mit aufgenommen.

3.4 Filterung automatisch generierter Tweets

Um die statistische Auswertung nicht zu verzerren, haben wir insgesamt 25.736 Tweets entfernt, die eindeutig automatisch generiert wurden. Dazu zählen z.B. Benachrichtigungen aus Videospielen, Musik-Updates oder Foursquare-Mitteilungen. Es wurden alle Tweets als ungültig markiert, die eines der Folgenden Tokens enthalten: '@YouTube', 'Gutschein', '#4sq', '#androidgames', '#nowplaying', '#np', 'Verkehrsmeldungen' und 'Wetterdaten'. In diesem Fall wurde das Attribut `valid` auf 0 (false) gesetzt.

4 Statistics

4.1 Dialog und Followerzahl

4.1.1 Verteilung der Followerzahl

Während des untersuchten Zeitraums zwischen dem 1. April und dem 30. April 2013 waren 1.577.083 unterschiedliche Accounts auf Twitter aktiv. Dabei zeigte sich zum einen eine extreme Ungleichverteilung hinsichtlich der Follower: Die obersten 5 % der User vereinen mit 4.549.895 99,9 % der Follower auf sich, während bereits die nächste Stufe der 90 - 95% follower-stärksten User im Schnitt nur noch 1.302 Follower aufweisen. Der Median betrug 126 Follower. Die Zahlen zeigen, dass das unterste Drittel der User seine Follower aus dem persönlichen Umfeld rekrutiert. Dieser Vermutung sollte sich natürlich auch in der Art und Weise einer dialogischen Kommunikation niederschlagen. So war es für uns

naheliegender, die letzten 5 Prozent der User nicht zu betrachten, da die extrem hohe Followerzahl von über 4 Mio Dialoge praktisch komplett ausschließt. Stattdessen scheint die vorletzte Kohorte mit durchschnittlich 1.302 Followern noch tatsächliche Dialoge (*real dialogs*) durchzuführen.

%	Follower	%	Follower	%	Follower	%	Follower
0-5	1	25-30	37	50-55	154	75-80	406
5-10	3	30-35	55	55-60	185	80-85	528
10-15	7	35-40	77	60-65	223	85-90	750
15-20	12	40-45	100	65-70	267	90-95	1.302
20-25	22	45-50	126	70-75	325	95-100	4.549.895

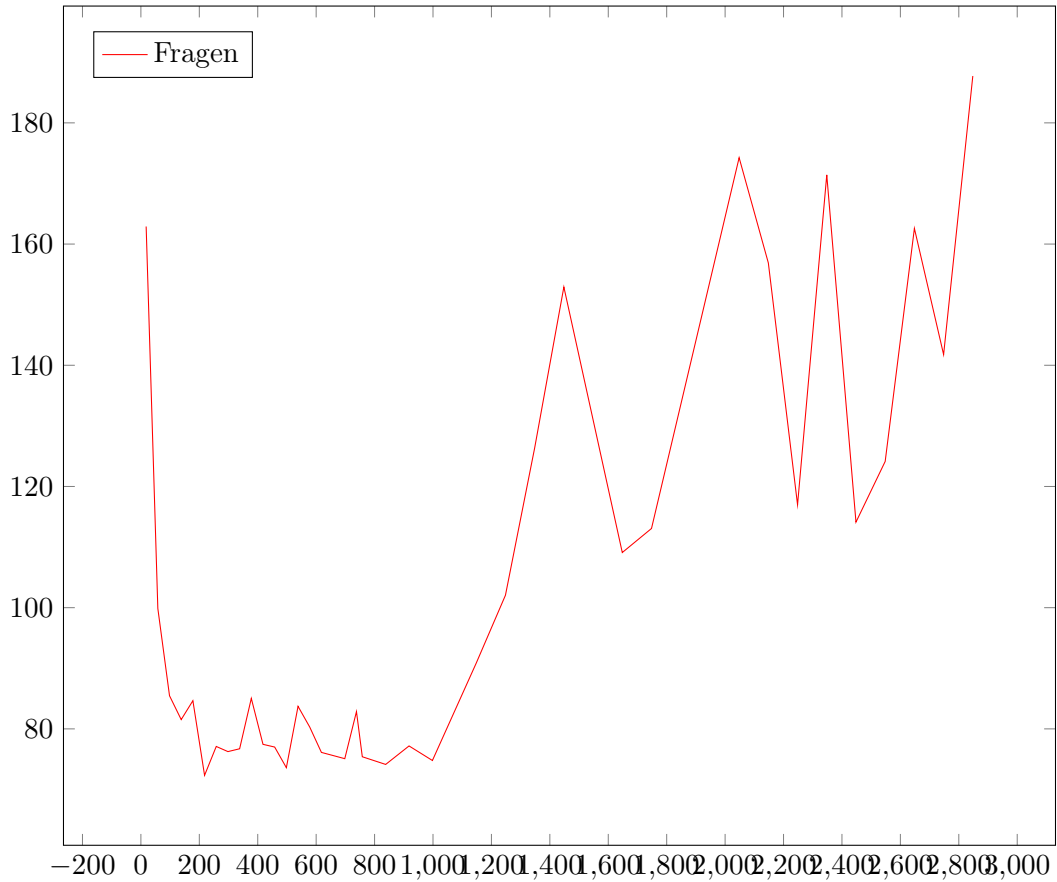
Tabelle 1: Followerzahl

4.2 Dialog und Followerzahl

4.2.1 Antwortzeit in Abhängigkeit von der Followerzahl

Ein Beispiel für unsere *real dialogs* - These ist die durchschnittliche Antwortzeit, die eine getweetete Frage erzielte. Wir setzten diese Zeit mit der Followerzahl des Users in Verbindung und erhielten eine Kurve, die darauf schließen lässt, das *real dialogs* nur bis zu einer Größenordnung von ca. 1.000 Followern stattfinden. Man kann erkennen, dass erwartungsgemäß die Zeit bis zur Beantwortung einer Frage bei sehr wenigen Followern hoch ist. Ab ca. 100 Followern jedoch pegelt sich die Beantwortungszeit einer Frage auf einen Wert um 80 Minuten ein. Ab einer Followerzahl von 1.000 schnellte die Zeit nach oben bzw. lässt keinen linearen Zusammenhang mit der Followerzahl erkennen. Dies verwundert, sollte doch bei einer hohen Zahl von Followern eine schnellere Beantwortung erwartet werden. Wir deuten die Zahlen als einen Hinweis darauf, dass in diesen Größenordnungen eben kein persönlicher Kontakt des Users zu seinen Followern vorhanden ist und die Follower einer aufgeworfenen Frage gegenüber indifferent reagieren, da einerseits ohnehin eine große Menge an Antworten vorhanden sein dürfte und sich der Follower zudem der Tatsache bewusst ist, sich eben in keinem *real dialog* zu befinden.

Länge von normalen Tweets und Fragen

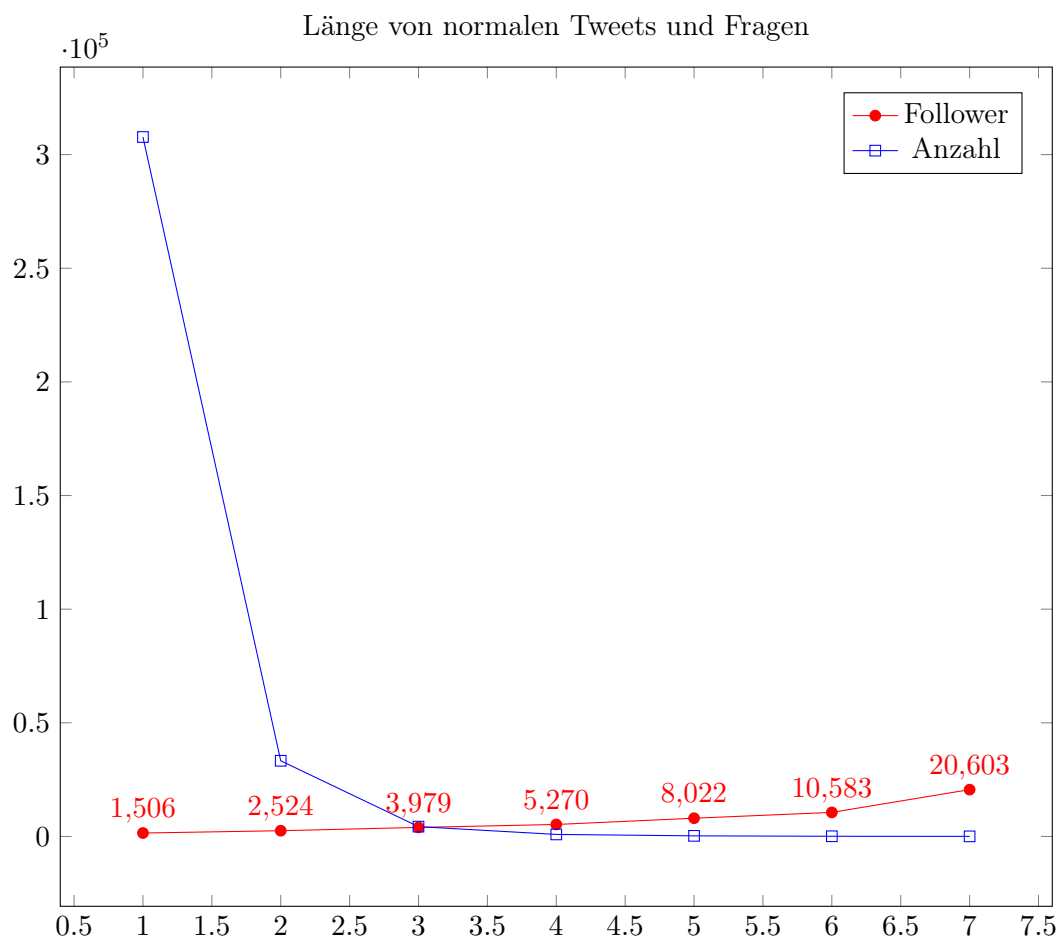


4.2.2 Breite des Dialogbaums in Abhängigkeit von der Followerzahl

Einen weiteren Hinweis darauf, ob es sich um einen *real dialog* handelt versprochen wir uns vom Blick auf die Breite des Dialogbaums. Unsere Überlegung war, dass der dialogische Charakter einer Twitter-Konversation durch einen „schlanken“ Baum eher abgebildet wird, als durch einen „breiten“, da Breite nur darauf hindeutet, dass eine Frage von vielen Followern gelesen wurde und entsprechend singulär beantwortet wird, d.h., ohne nochmals auf andere Antworten einzugehen. Der Blick auf die Daten bestätigte unsere Vermutung: Zunächst kann man ablesen, dass eine übergroße Zahl von Dialogen genau aus 2 Tweets bestehen - einem initialem Tweet und einer Antwort. Ebenso jedoch kann abgelesen werden, dass mit zunehmender Followerzahl eines Users die Dialogbäume breiter werden. Betont werden muss, dass hier nicht betrachtet wird, wie viele unterschiedliche Teilnehmer sich an einem Dialog beteiligen, sondern wie viele Follower der Dialog-Eröffner hat. Es bleibt jedoch bei der Feststellung, dass eine größere Followerzahl mit einer größeren Breite eines Dialogbaums einhergeht und somit ein Dialog gewissermaßen mäandert.

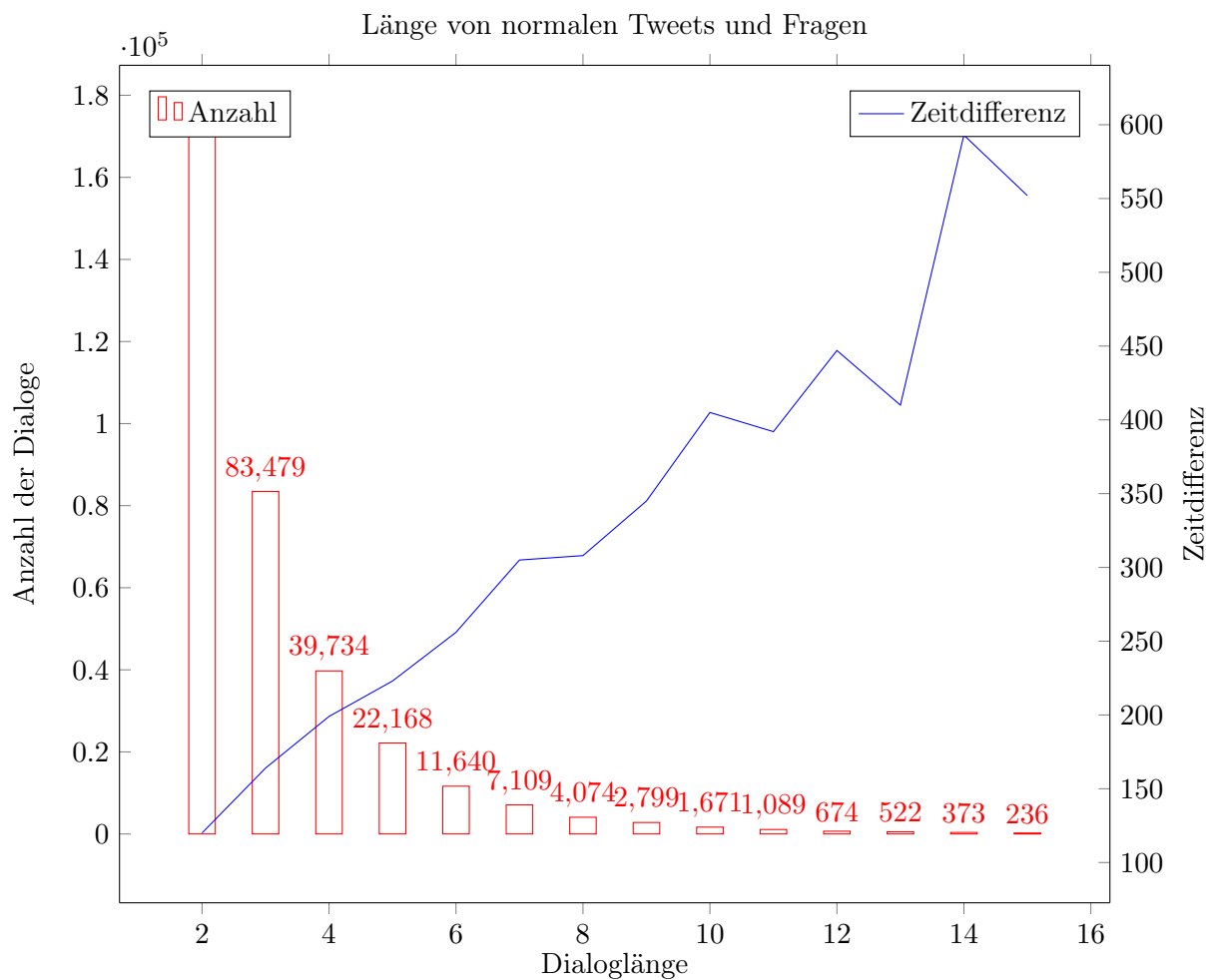
Breite des Dialogbaums	Durchschnittliche Follower	Anzahl
1	1.506	307.745
2	2.524	33.297
3	3.979	4.323
4	5.270	890
5	8.022	263
6	10.583	83
7	20.603	40

Tabelle 2: My caption



4.3 Dialoglänge und Dialogdauer

Twitter ist ein extrem schnelllebiges Medium. Beiträge anderer User können sehr schnell aus dem Blickfeld der Follower geraten. Damit läuft eine erwünschte Konversationseröffnung eines Users Gefahr, nicht wahrgenommen zu werden. Diese missglückten Versuche konnten von uns natürlich nicht nachgewiesen werden. Trotzdem konnten wir die Dauer von Konversation messen, indem wir die Differenz zwischen Dialogeröffnung und dem letzten Tweet des Dialoges zum Zeitpunkt der Erstellung des Korpus maßen. Die Dialoglänge selbst ermittelten wir, indem wir die Länge des längsten Pfades des Dialogbaums ermittelten. Man kann erkennen, dass die übergroße Zahl von Dialogen eine Lebensdauer von weniger als 3 Stunden hat und zudem kaum Dialoge stattfinden, die länger als einen Tag dauern.



Dialoglänge	Zeitdifferenz in Min	Anzahl Dialoge
2	120	170.298
3	164	83.479
4	199	39.734
5	223	22.168
6	256	11.640
7	305	7.109
8	308	4.074
9	345	2.799
10	405	1.671
11	392	1.089
12	447	674
13	410	522
14	593	373
15	552	236

Tabelle 3: My caption

4.4 Dialog und Fragen

Besonders interessiert hat uns, ob Twitter genutzt wird, um Fragen zu stellen und zu hoffen, diese beantwortet zu bekommen.

Eine zentrale Aufgabe stellte dabei die Identifizierung von Fragen dar. Wir gingen davon aus, dass aufgrund der Beschränkung auf 140 Zeichen viele Twitter-Nutzer auf die Angabe von Fragezeichen verzichten würden:

Wieso flüstern die so maaaaan (ID: 318495612649762816)

Daher untersuchten wir einen Tweet zusätzlich darauf, ob er sogenannte w-Wörter (*wer, wie, was, wodurch* ...) enthielt (satzinitial sowie satzintern) und ermittelten folgende Verteilung hinsichtlich der Attribute Fragezeichen und w-Wörter:

Fragezeichen	2.374.174
Tweet mit w-Wort	588.322
Tweet mit initialem w-Wort mit Fragezeichen	92.077
Tweet mit initialem w-Wort ohne Fragezeichen	51.465

Tabelle 4: My caption

Aus dieser Verteilung kann abgelesen werden, dass der potentielle Zugewinn durch das Betrachten von Tweets mit initialen w-Wörtern und ohne Fragezeichen gering wäre, zumal eine eindeutige Klassifizierung als Frage dann immer noch nicht gegeben wäre:

Wen du endlich erwachsen bist und mich in Ruhe lässt lass es mich wissen ansonsten viel spaß noch! Wirst schon sehen! (ID: 318501739030536192)

Da es bei einer so großen Datenmenge aus Geschwindigkeitsgründen nicht möglich ist, Fragen akkurat und linguistisch korrekt anhand syntaktischer Merkmale zu identifizieren, mussten wir uns einer Näherungslösung bedienen. Jeweils 100 zufällig gewählte Tweets, die entweder als wh-question, als Tweet mit Fragezeichen oder als Nicht-Frage getagged waren, überprüften wir händisch und ermittelten folgende Ergebnisse:

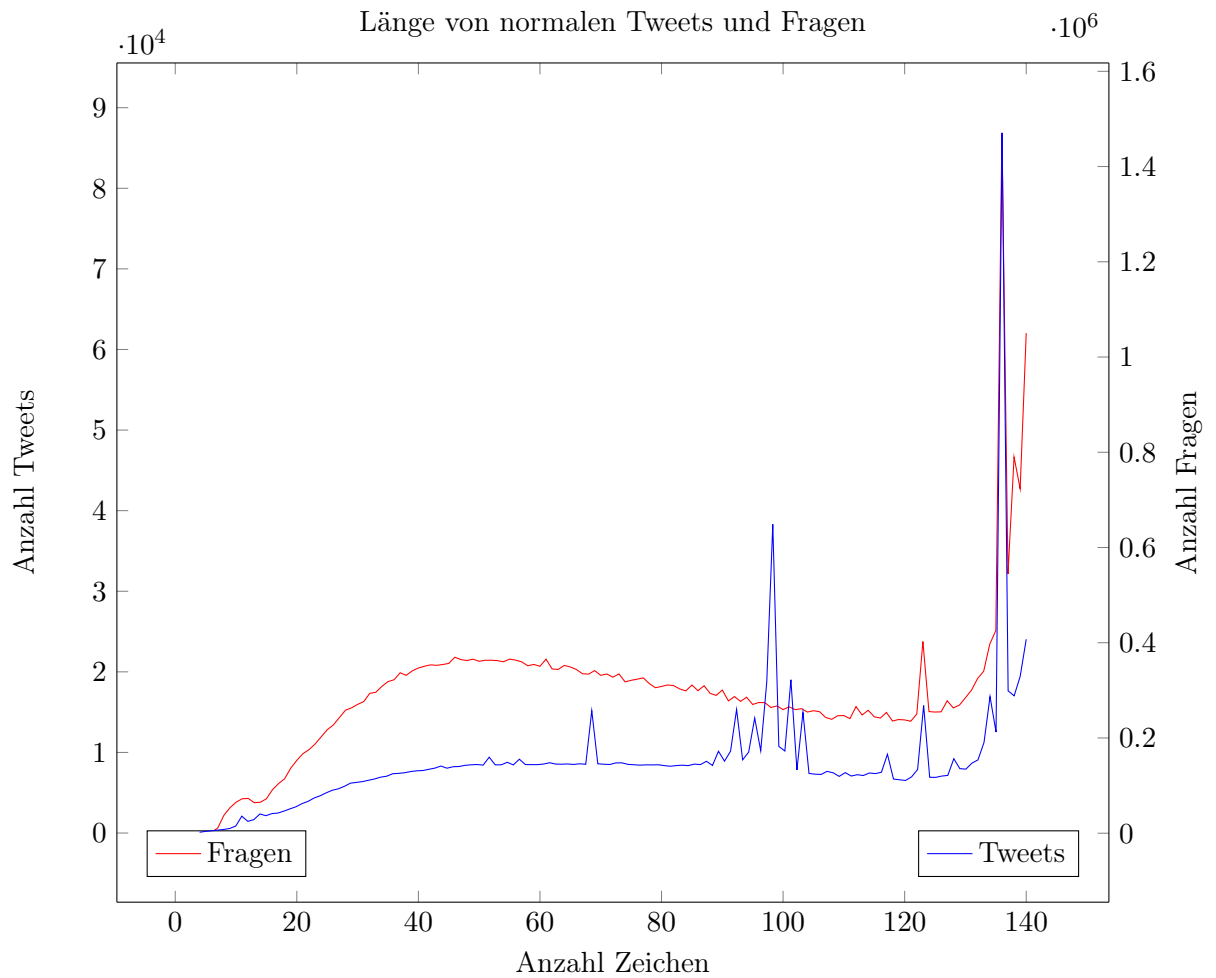
	wh_question=1	question_mark=1	is_question=0
richtig	73	100	100
falsch	27	0	0

Tabelle 5: My caption

Wir mussten erkennen, dass nicht nur der Anteil potentieller Fragen, bei denen auf ein Fragezeichen verzichtet wurde, gering war, sondern gleichzeitig auch die Überprüfung hinsichtlich der Existenz von w-Wörtern eine Fehlerquote von 27 % aufwies. Gleichzeitig war in unserer Stichprobe jeder der untersuchten Tweets mit Fragezeichen tatsächlich eine Frage. Das Ergebnis dieses Test widersprach unserer Erwartung, dass auf den Einsatz von Fragezeichen verzichtet würde. Das Attribut `question_mark` stellte sich als zuverlässigstes Indiz zur Fragenerkennung heraus. Wir beschränkten uns daher auf dieses Attribut, um Tweets als Fragen zu identifizieren.

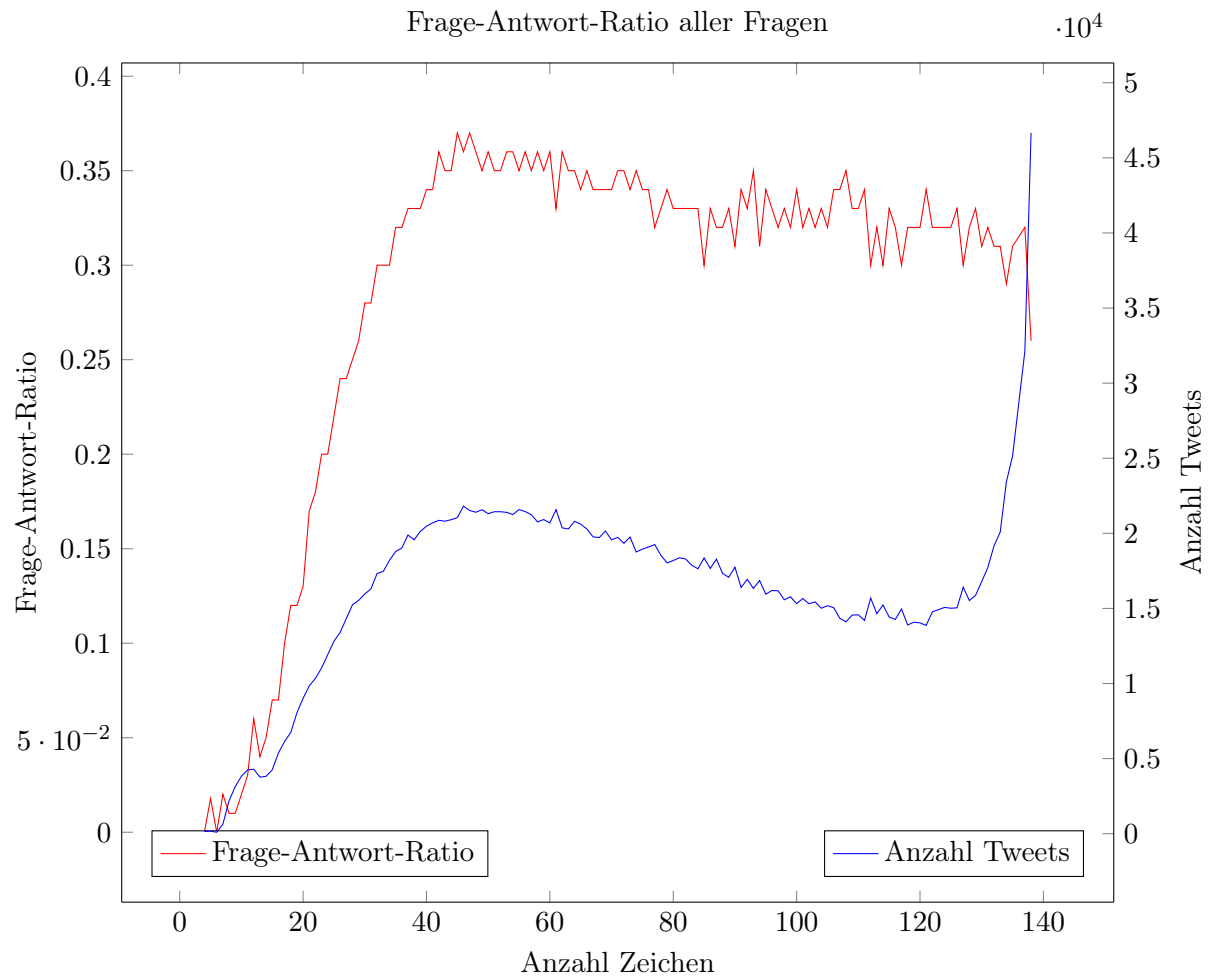
4.4.1 Länge von normalen Tweets und Fragen

Zunächst warfen wir einen Blick auf die durchschnittliche Tweetlänge. Unsere Vermutung war, dass es bei den Fragen einen höheren Anteil von Nachrichten mit geringer Länge geben würde, da oftmals einfache w-Wörter gesendet werden. Dieser Eindruck bestätigte sich jedoch nicht. Stattdessen folgen Fragen ebenso wie normale Tweets dem selben Muster, dass die maximal mögliche Zeichenanzahl ausgenutzt wird. Interessant ist die Tatsache, dass sowohl in Fragen als auch normalen Tweets ein Maximum an Tweets bei der Länge 136 gemessen wurde. Dies deutet darauf hin, dass User, die an die maximal zulässige Marke von 140 Zeichen kommen, ihren Tweet kürzen, indem lange Wörter durch kürzere ersetzt werden.



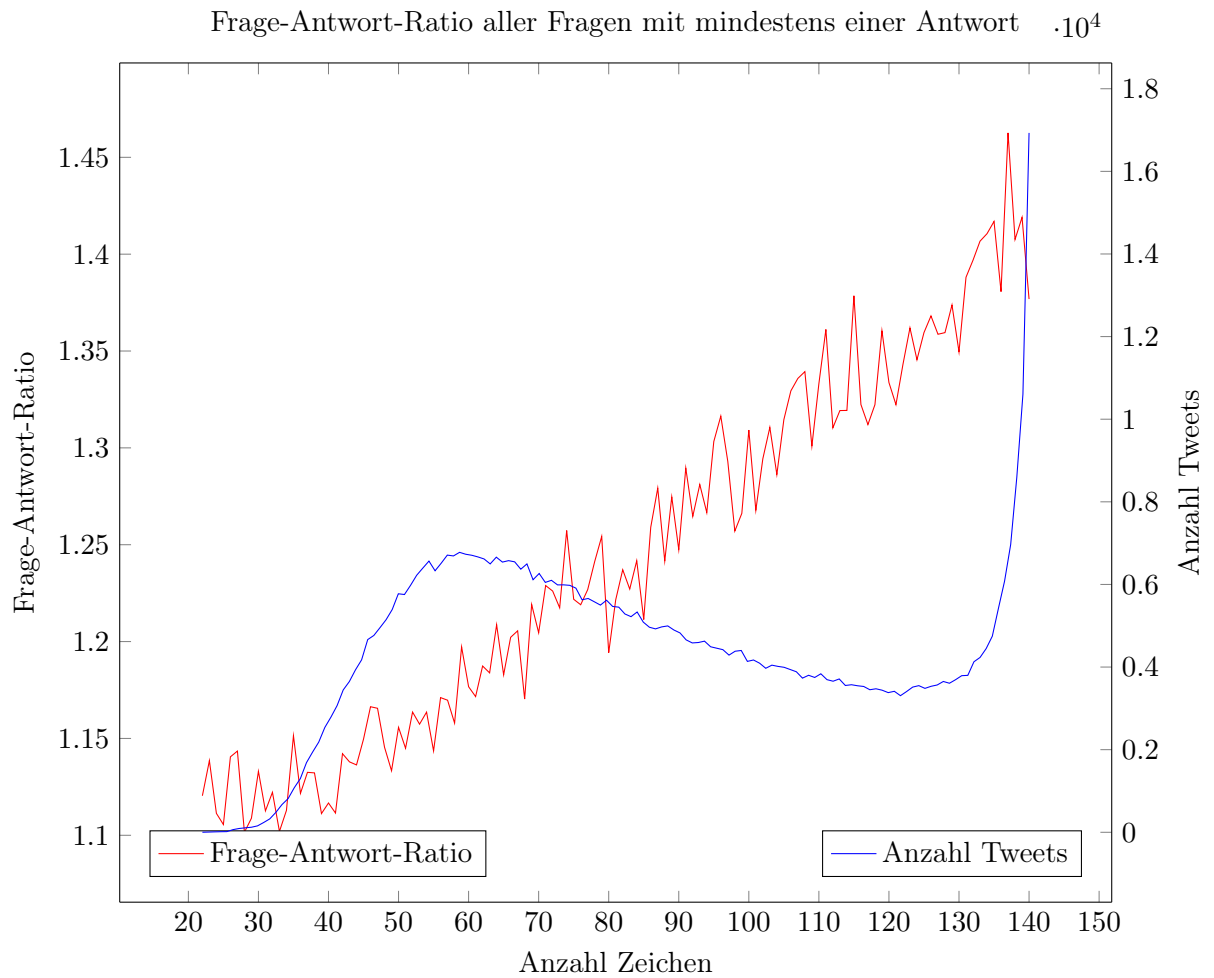
4.4.2 Frage-Antwort-Ratio

Wir wollten wissen, wie die Länge einer Frage die Chance erhöht, auch eine Antwort zu erhalten und nicht in der Tweetflut der Follower unterzugehen. Wir vermuteten, dass eine längere und damit klarer formulierte Frage mehr Antworten bzw. überhaupt eine Antwort erhalten sollte, als kürzere. Um diese Frage beantworten zu können, ermittelten wir eine *Frage-Antwort-Ratio* sämtlicher Fragen-initialer Dialoge mit jeweils der gleichen Länge der Frage. Dabei ermittelten wir somit für beantwortete Fragen schlicht die Zahl der Antworten. Eine Frage mit 3 Antworten hat somit eine *Frage – Antwort – Ratio* = 3. Da viele Fragen keine Antwort erhalten wird sich die *Frage – Antwort – Ratio* für jeweilige Zeichenlängen maximal 1 erreichen.



Man kann erkennen, dass die Frage-Antwort-Ratio bei sehr kurzen Fragen ohnehin sehr gering ist. Ab einer Zeichenlänge von 40 steigt sie spürbar an, lässt danach jedoch keine weitere Tendenz zum Anstieg erkennen. Dies spiegelt die erwähnte Gefahr wieder, dass eine Frage unbeantwortet bleibt.

In einem zweiten Schritt untersuchten wir die Frage-Antwort-Ratio bei allen den Fragen, die mindestens eine Antwort erhielten. Aus diesen Gründen ist die *Frage – Antwort – Ratio* ≥ 1 . Auffällig ist die kontinuierlich ansteigende Kurve der Frage-Antwort-Ratio mit zunehmender Länge der initialen Frage.

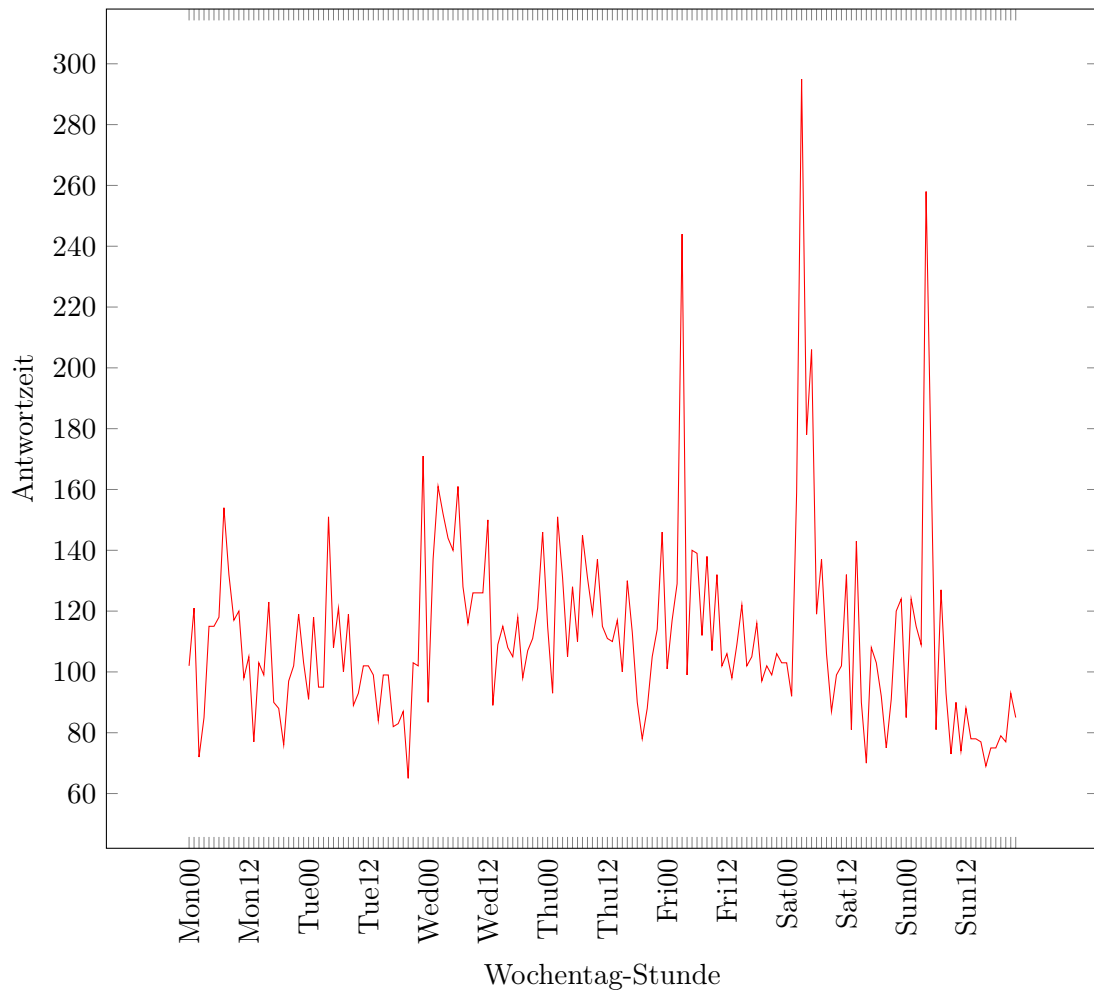


4.4.3 Fragen und Antworten im Tagesverlauf

Wir wollten wissen, ob man aus den erhobenen Daten Hinweise gewinnen kann, zu welcher Tageszeit das Stellen einer Frage auf Twitter besonders sinnvoll erscheint, wenn ein User tatsächlich an einer Antwort interessiert ist.

Zunächst interessierte uns, wie schnell ein User zu welcher Tageszeit im Durchschnitt eine Antwort auf seine Frage erhielt. Wir vermuteten, dass eine Nachts gesendete Frage viel länger auf eine Beantwortung warten müsse, als eine zur Prime-Time veröffentlichte Frage. Die Daten offenbarten folgende Muster:

Wartezeit zur ersten Antwort nach Wochentag-Stunde



Die durchschnittlich kürzeste Antwortzeit mit 64 Minuten entdeckten wir Dienstags um 21 Uhr. Es könnte also vermutet werden, dass die Abendstunden ein guter Zeitraum sind, um Fragen zu stellen und schnell beantwortet zu bekommen. Wenig verwunderlich finden sich die Maximalwerte bei den Antwortzeiten in den Nachtstunden. So können die 3 auffälligen Werte sämtlich in den sehr frühen Morgenstunden gefunden werden (Maximalwert 294 Minuten am Samstagmorgen 4 Uhr).

Neben der Geschwindigkeit, mit der eine Frage eine Antwort erhält gibt es natürlich noch eine zweite Kategorie, die für einen Fragensteller interessant ist: die Menge der Antworten auf eine Frage. Um diesen Aspekt zu beurteilen betrachteten wir die unterschiedlichen Wochentage über einen Zeitraum von einem Monat. Wir unterteilten jeden einzelnen Tag nochmals in seine 24 Stunden und ermittelten die Frage-Antwort-Ratio für diese Zeiträume.



Auch in diesem Fall sind die Einbrüche in den Nachtstunden auf den ersten Blick erkennbar. Ebenso erkennbar sind jedoch die immer wiederkehrenden Muster in den Abendstunden. Während in den Abendstunden nach 21 Uhr die Frage-Antwort-Ratio stets absinkt, erreicht sie gegen Mitternacht jeweils ihr Hoch. Dies gilt für sämtliche Wochentage. Man scheint also die Medienkonsum-Gewohnheiten der User entdecken zu können: Während der Abend gesellschaftliche Aktivitäten gehören scheint für User der Blick auf den Twitter-Account vor dem Zu-Bett-Gehen üblich zu sein. Die höchste Fragen-Antwort-Ratio wird Samstag Mitternacht erzielt.

5 Echofragen

Twitter ist ein junges und dynamisches Medium. Seine Entstehungsgeschichte spielt sich mit dem Jahr 2006 in einer Zeit ab, als bestimmte Termini der Internetsprache sich bereits verfestigt hatten. So ist die Verwendung des @-Symbols übernommen worden aus der Email-Domäne. Ein anderes Symbol – das #-Zeichen – ist von Twitter selbst prominent gemacht worden. Aus linguistischer Sicht interessant ist dabei, dass diese Symbole nicht dem bisher gebräuchlichem Kanon geschriebener Sprache angehörten. Stattdessen zeigt sich ein kreativer Umgang mit Sprache, der zum Einen den gewohnten soziolinguistischen Motivationen der User geschuldet ist, zum Anderen aber schlicht durch die technischen Anforderungen erzwungen werden. So führt die Begrenzung auf 140 Zeichen zu einer stetigen Fortentwicklung des Inventars an Abkürzungen. Der Einsatz von *Emoticons* führt Elemente von ikonischer Sprache in geschriebener Kommunikation ein, indem menschliche Mimik durch verfügbare Symbole dargestellt wird. Dieser kreative Umgang mit Sprache macht Twitter für die Linguistik zu einem interessanten Forschungsfeld, da hier Sprache niedergeschrieben vorgefunden wird, die deutlich von normierter Schriftsprache abweicht.

Diese auch unter dem Terminus Bestätigungsfrage, Vergewisserungsfrage Frageform dient vor allem einer rückwärtsgewandten Funktion, indem eine vorangegangene -eventuell missverstandene- Aussage mit fast identischem Wortmaterial wiederholt und in eine Frage umgewandelt wird um so nähere Erläuterungen zu erfragen. Echofragen sind ein Phänomen gesprochener Sprache. Oftmals werden sie verwendet, um akustische Verstehensprobleme auszuräumen. Dadurch gewinnt die Echofrage einen rein funktionalen Charakter - auf eine aufwendige Suche nach neuer Lexik wird vom Fragenden verzichtet. Stattdessen wird der Fokus der Nachfrage durch Intonation dargestellt. So ist im Beispiel die Tatsache, dass ein Pilot ein Flugzeug absichtlich zum Absturz bringt so außergewöhnlich, dass der Hörer dies noch einmal bestätigt wissen möchte.

- (3) Der Pilot hat das Flugzeug absichtlich abstürzen lassen!
Der Pilot hat das Flugzeug ABSICHTLICH abstürzen lassen?

Wir nutzen das bereits vorhandene Attribut `is_question` der einzelnen Tweets im Korpus, um potentielle Echofragen aufzuspüren. Für sämtliche Fragen ermittelten wir die Tweets, auf die durch die Frage geantwortet wurde. Durch eine einfache Ermittlung der *Kosinus-Ähnlichkeit* bestimmten wir die Ähnlichkeit zwischen Stimulus-Tweet und dem Fragen-Tweet, der sich anschloss. Wir setzten die Schwelle der Kosinus-Ähnlichkeit bewusst hoch, da das Prinzip Echofrage auf großer Textähnlichkeit beruht. Wir wählten eine Kosinus-Ähnlichkeit > 0.8 als Schwellwert.

Die einfachste Methode, auf Twitter eine Echofrage zu stellen, ist das Voranstellen oder Anfügen von Fragezeichen auf einen Initial-Tweet:

- (4) So langsam kann man bei Hummels dann auch nicht mehr von unabsichtlich sprechen ;-)

(ID: 329329842296348673)

????? "@Maverick75HH: So langsam kann man bei Hummels dann auch nicht mehr von unabsichtlich sprechen ;-)"

(ID: 329330095368056833)

- (5) Ich nenn das jetzt mal "leichten DM-Mitteilungsdrang":D

(ID: 322048566749167617)

"@mussmansein: Ich nenn das jetzt mal "leichten DM-Mitteilungsdrang":D" ???

(ID: 322048698060255232)

Auffällig ist, dass dabei bewusst die formellen Regeln der Schriftsprache überschritten werden und stattdessen durch -teilweise mehrfache- Duplikation von Satzzeichen der pragmatische Charakter unterstrichen wird. Dabei sind auch Mischungen von Satzzeichen möglich:

- (6) @ciffi Konstruktivisten nennen es auch gerne assimilieren #bfas13

(ID: 327702706808578048)

?!? "@gibro: @ciffi Konstruktivisten nennen es auch gerne assimilieren #bfas13"

(ID: 327703517647208448)

In vielen Fällen geben die User jedoch einen Hinweis auf den Focus ihrer Nachfrage. Im Gegensatz zur gesprochenen Sprache wird hier jedoch nicht Intonation verwendet. Das immer wiederkehrende Muster, dass wir beobachteten war das Voranstellen oder Anfügen der interessierenden lexikalischen Einheit.

- (7) #DSDS ist heute mal voll langweilig.

(ID: 320608166368903169)

heute? ;D @Binane1 #DSDS ist heute mal voll langweilig.

(ID: 320608518434611201)

- (8) @FitriNuraini_ Ich will zum ersten ya spielen :) !!

(ID: 327772778302431232)

spielen? RT @yazzin22: @FitriNuraini_ Ich will zum ersten ya spielen :) !!

(ID: 327773283502145536)

Ein Großteil der entdeckten Dialoge kann jedoch nicht als Echofragen identifiziert werden. Sie folgen mehrheitlich dem Muster, den initialen Tweet lediglich zu zitieren und eine neue Frage zum Inhalt zu stellen:

- (9) Sandmännchen beendet, jetzt die Kinder ins Bett und dann Feierabend bei einem Glas Wein... :-D

(ID: 318770276391321600)

*“@WeinKnueller: Sandmännchen beendet, jetzt die Kinder ins Bett und dann
Feierabend bei einem Glas Wein... :-D und welchen Wein gibt es ?
(ID: 318772980769501185)*

Echofragen können exemplarisch für die Hybris aus geschriebener und gesprochener Sprache bei Twitter stehen: Die technischen Möglichkeiten versetzen den User in die Lage, ebenso spontan und schnell wie in der gesprochenen Sprache und ohne zusätzliche Kreativität in die Formulierung zu investieren Nachfragen zu stellen und dabei den Focus gezielt zu lenken. Dabei werden Eigenschaften der gesprochenen Sprache wie Intonation ersetzt durch einfache Nennung der interessierenden lexikalischen Einheit. Bestehende formale Regeln werden dabei ignoriert.