

Aktualität und Gedächtnis

Zur korpusanalytischen

Untersuchung von

Gegenwartsliteratur auf Twitter

Meier-Vieracker, Simon

simon.meier-vieracker@tu-dresden.de
TU Dresden, Germany

Kreuzmair, Elias

elias.kreuzmair@uni-greifswald.de
Universität Greifswald, Germany

Twitter gilt als ein besonders präsentisches Medium. Mit der in der deutschen Fassung in der Eingabemaske gestellten Frage „Was gibt’s Neues?“ und überhaupt der Möglichkeit, sich gleichzeitig mit einem an eine potenziell unbegrenzte Öffentlichkeit gerichteten Instant Messenger in Echtzeit an öffentlichen Diskursen zu beteiligen, lädt Twitter in besonderer Weise zu gegenwartsorientiertem Schreiben ein. Auch aus einer Rezipierendenperspektive kann der primäre Zugriffsmodus der Timeline als Echtzeit-Nachrichtenticker beschrieben werden (Hermes 2021). Zugleich machen die medialen Affordanzen der Persistenz und der Durchsuchbarkeit (boyd 2014: 11) der auf Twitter platzierten Inhalte die Plattform zu einem umfassenden und weitgehend frei zugänglichen Archiv und mithin zu einem breit und vielfältig genutzten Medium des digitalen Gedächtnisses. Dass Twitter sich inzwischen selbst nicht nur als auf Aktualität gerichtetes Medium, sondern auch als Gedächtnismedium versteht, zeigt die Öffnung des Twitter-Archivs mit der neuen API für Forscher:innen (Torres/Trujillo 2021).

Dieses Spannungsverhältnis zwischen gegenwartsbezogenem Schreiben und digitalem Archiv wird nicht nur in zeitdiagnostischen Texten (Osten 2004, Renouard 2018), sondern auch im gegenwartsliterarischen Diskurs auf Twitter selbst immer wieder thematisch. Schriftsteller:innen nutzen Twitter – oder werden dort überhaupt erst zu solchen – und die damit einhergehenden Publikations- und Vernetzungsmöglichkeiten, wobei Kurztexte mit eigenem literarischem Wert, Metadiskurse über Literatur- und den Literaturbetrieb und schließlich auch interaktionsorientierte Kommunikation ineinander übergehen. Bei dieser manchmal als „Twitteratur“ bezeichneten Domäne (Kreuzmair 2016) handelt es sich also um Gegenwartsliteratur und -reflexion, die flüchtig und momentbezogen ist und sich dennoch gleichsam selbst archiviert. Durch die hohe Selbstreflexivität des Diskurses im Sinne ständiger Selbstbeobachtung werden dabei die Gegenwärtigkeit, aber auch Erinnerbarkeit des eigenen Schreibens immer wieder reflektiert.

Für das DFG-Projekt „Schreibweisen der Gegenwart. Zeitreflexion und literarische Verfahren nach der Digitalisierung“ (2020–2022), das nach Wechselwirkungen von Zeitreflexion und literarischen Verfahren unter den Bedingungen der Digitalisierung fragt, haben wir im Februar 2020 ein Twitter-Korpus mit den Timelines von 117 öffentlichen Accounts erstellt, die sich in einem erweiterten Sinne der deutschsprachigen Literaturszene zurechnen lassen. Der Datenerhebung war eine teilnehmende Beobachtung über zwei Monate vorausgegangen, die neben der Identifizierung relevanter Accounts auch ergeben hat, dass

eine starke Trennung Autor*innen/Literaturbetrieb nicht sinnvoll ist. Unter Nutzung der API über die Software rtweet (Kearney 2018) konnten den Beschränkungen der API entsprechend pro Account bis zu 3000 Tweets erhoben werden, das Korpus umfasst insgesamt 219.450 Tweets aus dem Zeitraum 2009–2020 im Umfang von 3.552.773 Wörtern. Für korpuslinguistische Untersuchungen wurden die Daten in einem XML-Format mit umfangreichen Metadaten aufbereitet, das über den Text hinaus auch interaktive Aspekte wie Reply-Strukturen (Hoppe et al. 2018) und Social Media-charakteristische Metadaten wie Anzahl der Likes und Retweets erfasst. Die Entscheidung für ein eigenes Datenmodell begründet sich durch die fehlenden Standards zur Encodierung von Social Media-Texten (etwa nach TEI), welche die für unsere Fragestellung relevante Interaktivität erfassen. Die Daten wurden mithilfe der auf Social Media-Daten trainierten Python-Module SoMaJo und SoMeWeTa (Proisl & Uhrig 2016; Proisl 2018) tokenisiert, nach Wortarten annotiert und lemmatisiert. Über die webbasierte Korpusanalyseplattform CQPweb (Hardie 2012), das äußerst flexible Abfragen der annotierten Daten und der Metadaten erlaubt, wird das Korpus den Projektbeteiligten zur Verfügung gestellt. Formulierungsmuster, semantische Profile, aber auch Interaktions- und Vernetzungsstrukturen können so computergestützt untersucht werden. Dafür können die in die Software implementierten korpuslinguistischen Methoden wie Keywords in Context, Kollokationsanalysen, Ngramm-Analysen, Distributionsanalysen und Keyword-Berechnungen genutzt werden. Darüber hinaus sind im Projekt auch andere digitale Textanalysemethoden wie etwa Topic Modeling zum Einsatz gekommen (Schöch 2017).

Das so erstellte und als Grundlage für die weiteren Analysen herangezogene Korpus ist somit eine Momentaufnahme, eine Fixierung und mithin auch ein Stillstellungs-Artefakt (Jäger 2011: 315) eines eigentlich äußerst fluiden Diskurses. Er wird so der auch quantifizierbaren und reproduzierbaren Analyse zugänglich, gleichsam als digital prozessiertes Datenerbe, und büßt dadurch aber zugleich jene Offenheit und Flüchtigkeit ein, die Social Media-Kommunikation in besonderem Maße auszeichnet.

Auf dem Poster stellen wir zum einen unsere Pipeline für die Korpuserstellung und -aufbereitung sowie unser Datenmodell für die Korpusrepräsentation vor. Zum anderen präsentieren wir exemplarische Analyseergebnisse zu Zeitreflexionen auf Twitter wie etwa die auffallende Häufigkeit der Trigramme „das erste Mal“ und „den ganzen Tag“, die einerseits die Zeitbezogenheit des Schreibens auf Twitter anzeigt und andererseits seine tagebuchartige Funktion als Archiv für bemerkenswerte Ereignisse vor Augen führt. Davon ausgehend werden wir die methodologischen Schwierigkeiten diskutieren, die sich aus der analytisch unumgänglichen Fixierung eines eigentlich fluiden Medienformates ergeben.

Da der Urheberrechtsstatus von Tweets bislang ungeklärt ist, ist eine vollständige Publikation des Datensatzes leider nicht möglich. Interessierten Forschenden Zugriff auf das Korpus in der verwendeten Analyseplattform CQPweb gewährt werden, wo im Korpus recherchiert werden kann.

Bibliographie

boyd, danah (2014): *It’s complicated: the social lives of networked teens*. New Haven: Yale University Press.

Hardie, Andrew (2012): “CQPweb — combining power, flexibility and usability in a corpus analysis tool”, in: *International Journal of Corpus Linguistics* 17(3): 380–409. 10.1075/ijcl.17.3.04har.

Hermes, Jürgen (2021): "Chirpy Humanities". Billet *Public Humanities* <https://publicdh.hypotheses.org/42> (letzter Zugriff 14.07.2021).

Hoppe, Imke, Lörcher, Ines / Neverla, Irene / Kießling, Bastian (2018): „Gespräch zwischen vielen oder Monologe von einzelnen? Das Konzept „Interaktivität“ und seine Eignung für die inhaltsanalytische Erfassung der Komplexität von Online-Kommentaren“, in: *Neue Komplexitäten für Kommunikationsforschung und Medienanalyse: Analytische Zugänge und empirische Studien (Digital Communication Research Band 4)*. doi:10.17174/dcr.v4.9.

Jäger, Ludwig (2011): „Intermedialität – Intramedialität – Transkriptivität. Überlegungen zu einigen Prinzipien der kulturellen Semiosis“, in: Arnulf Deppermann / Angelika Linke (eds.), *Sprache intermedial. Stimme und Schrift, Bild und Ton*, 301–323. Berlin, Boston: De Gruyter. 10.1515/9783110223613.299.

Kearney, Michael W. (2018): *rtweet: Collecting Twitter data*. Zenodo. 10.5281/zenodo.2528481.

Kreuzmair, Elias (2016): „Was war Twitteratur?“, in: *Merkur*. <https://www.merkur-zeitschrift.de/2016/02/04/was-war-twitteratur/> (letzter Zugriff 10.06.2021).

Osten, Manfred (2004): *Das geraubte Gedächtnis: Digitale Systeme und die Zerstörung der Erinnerungskultur. Eine kleine Geschichte des Vergessens*. Frankfurt am Main: Insel.

Proisl, Thomas (2018): „SoMeWeTa: A Part-of-Speech Tagger for German Social Media and Web Texts“, in: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* 665–670.

Proisl, Thomas / Uhrig, Peter (2016): “SoMaJo: State-of-the-art tokenization for German web and social media texts”, in: *Proceedings of the 10th Web as Corpus Workshop*. Berlin: Association for Computational Linguistics, 57–62. 10.18653/v1/W16-2607.

Renouard, Maël (2018): *Fragmente eines unendlichen Gedächtnisses*. Zürich: Diaphanes.

Schöch, Christof (2017): “Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama”, in: *Digital Humanities Quarterly* 11(2).

Tornes, Adam / Trujillo, Leanne (2021): *Enabling the future of academic research with the Twitter API*. https://blog.twitter.com/developer/en_us/topics/tools/2021/enabling-the-future-of-academic-research-with-the-twitter-api (letzter Zugriff 15.07.2021).