

Daniel Pfurtscheller (Wien)

Öffentlichen Sprachgebrauch auf Facebook untersuchen

Zugänge, Probleme, Erste Hilfe

1 Problemaufriss

Für den öffentlichen Sprachgebrauch im Internet ist Facebook, das mit 15 Jahren zur älteren Generation von Social-Media-Sites zählt, nach wie vor hochrelevant. Im deutschsprachigen Raum ist es die am meisten genutzte Social-Media-Plattform (Newman et al. 2019). Zu den Diensten gehören unter anderem Facebook-Seiten (Pages), die von Unternehmen, Parteien, Medien und anderen Institutionen oder Individuen betrieben werden und als öffentliche Angebote prinzipiell auch von nicht bei Facebook angemeldeten Personen eingesehen werden können. Solche öffentlichen Facebook-Seiten sind als *sites of engagement* zwischen gesellschaftlichen Institutionen und Individuen reichhaltige Quellen für die linguistische Forschung.

Im Vergleich zu anderen Plattformen bietet Facebook aber nur einen eingeschränkten Zugriff auf diese öffentlichen Sprach- und Interaktionsdaten (Freelon 2018). Während beispielsweise für Twitter viele Tools zur Datensammlung existieren und auch die Plattform selbst eine ausgebaute Suchmaske bietet, erschweren die limitierten Suchmöglichkeiten der Facebook-Plattform und das fehlende Angebot von einfach nutzbarer Software linguistische Projekte in Forschung und Lehre. Gleichzeitig stellen sich neben den praktischen Fragen an vielen Stellen auch forschungsethische Fragen im Umgang mit Onlinedaten.

2 Datenfangen auf Facebook: UI- und API-basierter Zugang

Für die Sammlung von Sprach- und Interaktionsdaten auf Social-Media-Plattformen gibt es zwei Möglichkeiten: Man kann Daten zum einen manuell auf der Oberfläche von *Facebook.com* oder der mobilen App sammeln und extrahieren. Dabei bewegt man sich im Rahmen des User Interface (UI) der Plattform (*screen-based data collection*, Androutsopoulos 2018). Zum anderen besteht die Möglichkeit,

<https://doi.org/10.1515/9783110679885-020>

die Programmierschnittstelle (Application Programming Interface, kurz API) von Facebook zu nutzen, mit der sich große Mengen von Sprach- und Metadaten maschinell sammeln und speichern lassen.

Beide Zugänge sind kombinierbar und haben bestimmte Vor- und Nachteile (siehe Tab. 1). Eine UI-basierte Datensammlung ist mit geringem technischen Aufwand allerdings nur für eine überschaubare Datenmenge durchführbar. Man erhält in der Regel menschenlesbare Daten, die für die maschinelle Auswertung erst weiter aufbereitet werden müssen. Bei einer API-basierten Datensammlung ist es genau umgekehrt. API-basierte Zugänge sind beliebig skalier- und automatisierbar, aufgrund von Änderungen seitens Facebook allerdings auch recht fragil (Freelon 2018).

Tab. 1: Ein Vergleich des UI-basierten und API-basierten Paradigmas

Datensammlung	UI-basiert	API-basiert
Modell	Benutzeroberfläche	Programmierschnittstelle
Datenrepräsentation	Menschenlesbarkeit	Maschinenlesbarkeit
Skalierbarkeit	gering	hoch
technischer Aufwand	gering	hoch
Automatisierbarkeit	gering	hoch
Robustheit	hoch	gering

Öffentliche Sprachverwendung auf Facebook wird durch Algorithmen, Affordanzen und das visuelle Design der Plattform vorstrukturiert (Androutsopoulos 2016). Explorationen von API-Daten haben in diesem Zusammenhang den Vorteil, gänzlich neue, quer zum vorgegebenen Design liegende Sichten auf die Daten zu ermöglichen (eine Plattformlogik ist aber freilich ist auch in API-Daten eingeschrieben). Um Aspekte der Multimodalität der Social-Media-Interaktion einzufangen, eignen sich UI-basierte Daten, die in Plattformdesigns kontextualisiert vorliegen. Die Kombination der Zugänge erlaubt eine qualitativ-heuristische Entdeckungsmethodologie, bei der sich erweiterte Suchmöglichkeiten und empirische Belegkraft mit multimodaler Detailanalyse verbinden lassen.

3 Facebook-spezifische Softwaretools und Extraktionspraktiken

Die Auswahl an Facebook-spezifischen Tools zur Datenextraktion und -analyse ist vergleichsweise gering. Der Schwerpunkt dieser Zusammenstellung liegt auf frei nutzbarer und offener Software.

Eine UI-basierte Datensammlung auf Facebook lässt sich durch Browser-Erweiterungen unterstützen: Ein Bookmarklet hilft dabei, die Kommentarverläufe unter Beiträgen ausklappen und automatisch erweitern zu lassen (Farley 2019). Für den Firefox-Browser gibt es eine eigene Screenshot-Erweiterung, die auch Möglichkeiten zur Anonymisierung bietet (Młokosiewicz 2019). Als Alternative zu Screenshots als Bilddatei bieten sich PDFs (über die Druckfunktion) oder HTML-Archive (Lormeau 2019) an, bei denen Texte durchsuchbar und alle Verlinkungen bestehen bleiben. Neben dem Sichern von einzelnen Postings per Copy-and-Paste oder als Screenshot lassen sich so auch ganze Kommentarverläufe in der ursprünglichen Präsentationsform dokumentieren und beispielsweise in qualitativer Analysesoftware (z.B. in *MAXQDA* oder *ATLAS.ti*) weiter auswerten.

Für eine API-basierte Datensammlung gibt es im Fall von Facebook nur wenig Anwendungssoftware. Zuletzt hast sich die Lage noch verschärft: Das webbasierte Tool *netvizz* (Rieder 2013), das in vielen Studien zur Datensammlung verwendet wurde, ist seit August 2019 nicht mehr nutzbar. Auch die Software *Facepager* (Jünger/Keyling 2018) ist von einer Drosselung des öffentlichen API-Zugriffs betroffen. Durch diese Einschränkungen werden Aktivitäten auf Facebook für die Forschung immer undurchsichtiger. Es bleibt abzuwarten, wie mit dieser Datenzugangslücke (Walker/Marcea/Basos 2019) umzugehen ist und welche methodischen Lösungen sich im Post-API-Zeitalter (Perriam/Birkbak/Freeman 2019) durchsetzen können.

4 Forschungsethische Herausforderungen

Neben technisch-praktischen Herausforderungen im Umgang mit Onlinedaten und ihrer Dynamik und Ephemeralität stellen sich in der Praxis auch forschungsethische Fragen. Die Association of Internet Researchers (AoIR) setzt sich kontinuierlich dafür ein, dass die Forschung im und über das Internet auf ethische und professionelle Weise durchgeführt wird und bietet Guidelines (Markham/Buchanan 2002). Im Rahmen einer Online-Ethnografie lassen sich UI-Daten auch individuell durch (im informierten Einverständnis abgesicherte) nutzerbasierte Daten ergänzen.

zen. Insbesondere für multimodale Analysen stellen sich dabei aber auch spezifische Fragen, die von bestehenden, zum Teil recht allgemeinen Guidelines nur in Ansätzen abgedeckt werden (Venema/Pfurtscheller/Lobinger i. Ersch.).

Literatur

- Androutopoulos, Jannis (2016): Mediatisierte Praktiken: Zur Rekontextualisierung von Anschlusskommunikation in den Sozialen Medien. In: Deppermann, Arnulf/Feilke, Helmut/Linke, Angelika (Hg.): Sprachliche und kommunikative Praktiken. (= Jahrbuch des Instituts für Deutsche Sprache 2015). Berlin/Boston, S. 337–367.
- Androutopoulos, Jannis (2018): Online data collection. In: Mallinson, Christine/Childs, Becky/Van Herk, Gerard (Hg.): Data collection in Sociolinguistics: Methods and applications. 2. Aufl. New York, S. 233–243.
- Farley, Jens-Ingo (2019): Expand All. JavaScript bookmarklet that expands all comments and replies in Facebook posts. <http://com.hemiola.com/bookmarklet/> (Stand: 7.8.2019).
- Freelon, Deen (2018): Computational research in the post-API age. In: Political Communication 35, S. 665–668.
- Jünger, Jakob/Keyling, Till (2018): Facepager. An application for generic data retrieval through APIs. <https://github.com/strohne/Facepager/> (Stand: 7.8.2019).
- Lormeau, Gildas (2019): SingleFile. A web extension to save a complete web page as a single HTML file. <https://github.com/gildas-lormeau/SingleFile> (Stand: 7.8.2019).
- Markham, Annette/Buchanan, Elizabeth (2002): Ethical decision-making and internet research: Recommendations from the AoIR Ethics Working Committee (Version 2.0). <https://aoir.org/reports/ethics2.pdf> (Stand: 7.8.2019).
- Młokosiewicz, Jakub (2019): Facebook Post Screenshot. Firefox web extension to save Facebook posts as images. <https://github.com/hckr/fb-post-screenshot/> (Stand: 7.8.2019).
- Newman, Nic/Fletcher, Richard/Kalogeropoulos, Antonis/Nielsen, Rasmus Kleis (2019): Reuters Institute Digital News Report 2019. Oxford. http://reutersinstitute.politics.ox.ac.uk/sites/default/files/inline-files/DNR_2019_FINAL.pdf (Stand 8.11.2019).
- Perriam, Jessamy/Birnbak, Andreas/Freeman, Andy (2019): Digital methods in a post-API environment. In: International Journal of Social Research Methodology. DOI: 10.1080/13645579.2019.1682840.
- Rieder, Bernhard (2013): Studying Facebook via data extraction: The Netvizz application. In: Proceedings of the 5th Annual ACM Web Science Conference. New York, S. 346–355.
- Venema, Rebecca/Pfurtscheller, Daniel/Lobinger, Katharin (i. Ersch.): Doing visual analysis online. Forschungsethische Herausforderungen und Handlungsempfehlungen zur Analyse vernetzter Bilder. In: Brantner, Cornelia (Hg.): Vernetzte Bilder. Köln.
- Walker, Shawn/Marcea, Dan/Bastos, Marco (2019): The disinformation landscape and the lockdown of social platforms. In: Information, Communication & Society 22, 11, S. 1531–1543.