

KONSTANZE MARX

DIE INSTAB-FORMEL

Ein Vorschlag für die Erstellung von Instagram-Datensammlungen für studentische Arbeiten

Abstracts: Als eine der beliebtesten Social-Media-Plattformen rückt Instagram immer mehr in den Fokus von Sprachwissenschaftler*innen. Bislang gibt es jedoch noch keine Standards für Korpora, die multimodale und ephemere Daten enthalten, wie sie für Instagram typisch sind. Gleichzeitig besteht auch bei Studierenden ein großes Interesse an einer wissenschaftlichen Untersuchung der Plattform, weshalb sich die Integration in die akademische Lehre aufdrängt. Datengrundlagen für Untersuchungen im Rahmen von Seminar- und Abschlussarbeiten entstehen häufig nach individueller Absprache. Mit der INSTAB-Formel wird hier ein Vorschlag unterbreitet, wie Schritt für Schritt eine Datensammlung angelegt werden kann, ohne dass Studierende über Programmierkenntnisse oder besondere technische Voraussetzungen verfügen. Die so generierte Datenbasis ermöglicht es Studierenden, qualitativ ausgerichteten Fragestellungen nachzugehen.

As one of the most popular social media platforms, Instagram is increasingly becoming interesting for linguistic studies. So far, however, there are no standards for corpora containing multimodal and ephemeral data, which are typical for Instagram. At the same time, there is also a great interest among students in a scientific investigation of the platform, which is why integration into academic teaching is an obvious option. Data bases for investigations in the context of seminar papers and theses are often created by individual agreement. With the INSTAB formular I make a proposal how a data collection can be created step by step without students having programming knowledge or special technical requirements. The database generated in this way enables students to pursue qualitatively oriented research questions.

Keywords: Soziale Medien, Instagram, Multimodalität, Korpusgenerierung, Ephemere Daten, Annotation, Transkription, Digitale Ethik

1. Multimodale Daten und heterogene Prozesse

Das Soziale Netzwerk Instagram gelangt zunehmend in den Interessensfokus Studierender, wenn sie sich, etwa im Rahmen einer Seminar- oder Abschlussarbeit, eingehender mit spezifischen internetlinguistischen Fragestellungen beschäftigen wollen.¹ Angesichts der Nutzungsprofile der App ist das nicht verwunderlich: Weltweit sind 1,48 Mrd. Nutzer*innen auf Instagram aktiv, die meisten (74%)² davon sind im Alter von 14 bis 34, eine Altersspanne, in der sich auch Studierende mehrheitlich befinden. Dass hier elizitierte Daten also durch eine große Nähe zur Lebenswelt von Studierenden gekennzeichnet sind, ist eine gute Voraussetzung für die Implementierung forschungsorientierten Lehrens und Lernens anhand von Sozialen Medien.

¹ Mein Dank geht an Josefine Guderian für eine studentische Perspektive auf den Text und Unterstützung bei der Transkription und Formatierung.

² <https://de.statista.com/statistik/daten/studie/1247275/umfrage/anteil-der-instagram-nutzer-nach-altersgruppen-und-geschlecht-weltweit/> (Stand: 2.4.2023).

Nun ist Instagram eine Plattform, die ihren Nutzer*innen ein breites Spektrum an Möglichkeiten zur Erzeugung von Inhalten zur Verfügung stellt: Beiträge, Reels, Stories – und damit statische und bewegte Bilder – sowie Gestaltungsoptionen, die unterschiedliche Modi (z.B. Boomerang), Filter, Effekte, Sticker (mit Umfragen etc.) oder Emojis integrieren. Auf diese Weise entsteht sehr komplexes, multimodales (teilweise ephemeres) Datenmaterial, das eine Herausforderung für die Generierung von Korpora darstellt. Diese schwierigen Bedingungen führen z.B. dazu, dass bei Untersuchungen einzelne oder mehrere Konstituenten der Kommunikate ausgeblendet werden. So wird in einigen Arbeiten der Fokus einzig auf Hashtags (Knierim/Achmann/Wolf 2022; Heyen 2022), auf „klassische Instagram-Beiträge mit statischem Bild“ (Schiefer 2022, S. 2), auf die Sprachebene, ohne (Bewegt)-Bilder (Gür-Şeker 2021b, S. 68), auf Kommentare (Vargas et al. 2021), Beiträge (Krammer 2021) oder einzelne Accounts (Issel-Dombert 2020) gerichtet. Gleichzeitig ist ein sehr hoher zeitlicher Aufwand für die Datenaufbereitung dokumentiert, vgl. Vargas et al. (2021, S. 3): „The entire process of annotation training took approximately 6 months“. Damit wäre die Bearbeitungszeit für eine Bachelor- oder Masterarbeit bereits ausgefüllt.

Erschwerend kommt hinzu, dass es bislang weder ein zugängliches Social-Media-Korpus gibt, das ephemere Kommunikate³ integriert (vgl. Beißwenger/Lüngen 2022) noch einheitliche Standards für die Transkription multimodaler Daten (u. a. Pfurt-scheller 2022). Der Anspruch an transkribierte Daten ist hingegen seit langem klar formuliert: Es braucht eine hinreichend gut definierte Darstellungsform, die reproduzierbare Analysen ermöglicht (Bateman 2008, S. 17). In den letzten Jahren lässt sich jedoch eher ein Trend hin zu dezentrierten Datensammlungen beobachten. Damit sind Korpora gemeint, die ad hoc, fragengeleitet und vor allem lokal – also unzugänglich für andere Forscher*innen – entstehen. Gerade für Untersuchungen zu ephemeren Modi, eben Reels, Stories oder auch Clips, die nicht nur für Instagram typisch sind, sondern z.B. auch für Snapchat oder TikTok, scheint das besonders praktikabel und verbreitet. Diese Entwicklung für die akademische Lehre fruchtbar zu machen, scheint also nicht abwegig.

Allerdings werden die Vorgehensweisen bei der Erstellung dieser Datengrundlagen kaum transparent dokumentiert. In Publikationen zu Social-Media-Plattformen, wie z.B. Instagram, sind häufig keine so präzisen Angaben zu finden, dass die Datengrundlage rekonstruiert oder zumindest die Vorgehensweise repliziert werden kann. Forscher*innen beschreiben den Prozess der Korpusgenerierung z.B. als Datenextraktion (Gür-Şeker 2021a, S. 69) oder folgendermaßen:

- das [...] Korpus von Instagram-Postings wurde *manuell* zwischen April und Juli 2020 *zusammengetragen* (Heyen 2022, S. 136)

³ Ephemere Kommunikate sind fluide und vergänglich, wie etwa Stories bei Instagram, die nach 24 Stunden gelöscht werden.

- „wurden alle [...] Instagram-Postings der sechs Lebensmittelunternehmen *gesammelt*, die [...] digital veröffentlicht wurden.“ (Krammer 2021, S. 196)
- „wurden die Postings des Accounts [...] *erhoben* und offline *gesichert*“ (Issel-Dombert 2020, S. 547)
- „We access Instagram *using a custom script* to emulate a user accessing the platform [...] *we created* a plugin for the chrome browser“ (Achmann et al. 2022, S. 1)
- The dataset was collected by *scraping* data from the social network Instagram (Vargas et al. 2021, S. 3) (Hervorhebungen jeweils von mir)

Welche Vorgänge und Prozesse genau mit dem hier genannten *manuellen Zusammentragen, Sammeln, Erheben* und *Sichern* verbunden sind, bleibt jedoch eine Black Box. Anhand der Rezeption von thematisch einschlägigen Studien können sich Studierende die Vorgehensweise also kaum erschließen. Auch das erwähnte Schreiben eines *Scripts* oder die *Erstellung eines Browser-Plugins* kann von Studierenden nicht erwartet werden. Zwar gibt es bereits in einigen sprachwissenschaftlichen Studiengängen entsprechende curriculare Angebote, Standard sind sie jedoch weder im Bachelor-, Master- noch Lehramtsstudium. Mit Blick auf die eng getakteten Studienpläne und die dadurch begrenzten zeitlichen Ressourcen kann nicht vorausgesetzt werden, dass Studierende bereit sind, Zusatzangebote wahrzunehmen oder sich autodidaktisch z.B. mit dem jüngst erschienenen Lehr- und Arbeitsbuch zur Python-Programmierung für Germanist*innen (Weißer 2022) fortzubilden. Auch von Dozierenden in den Geisteswissenschaften kann nicht erwartet werden, dass sie über so fundierte Programmierkenntnisse verfügen, dass sie bei auftretenden Problemen z.B. auch beim automatischen toolgestützten *Scrapen* helfen können. Zwar wird für die Erhebung von Instagram-Daten der Crawler *Instaloader* (ggf. mit DMI)⁴ empfohlen,⁵ beim Einsatz in der akademischen Lehre berichteten Studierende jedoch von Schwierigkeiten, sei es im Hinblick auf die eigentlich nur zu kopierende Python-Syntax oder die Kompatibilität mit dem eigenen Betriebssystem. Schon der misslingende Installationsversuch auf dem eigenen Rechner (so vorhanden) kann eine Hürde sein.

Insgesamt sind das keine guten Voraussetzungen für die Implementierung in die akademische Lehre, denn es ist nicht möglich, Studierenden auf Basis bereits vorhandener Studien einen Leitfaden für den Aufbau ihrer eigenen Datensammlung an die Hand zu geben.

Von dieser suboptimalen Ausgangssituation lassen sich aber weder Dozierende noch Studierende abhalten und so ist es gängige Praxis, dass analog zu den individuell erstellten Datensammlungen in größeren Forschungskontexten auch individu-

⁴ <https://github.com/digitalmethodsinitiative/dmi-instascraper> (Stand: 2.4.2023).

⁵ Vgl. das Lehrvideo von Simon Meier-Vieracker (TU Dresden) zur Erstellung von Social-Media-Korpora: www.youtube.com/watch?v=0Gy-1uICmm0 (Stand: 2.4.2023).

elle Lösungen für kleinere Projekte (Seminar- oder Abschlussarbeiten) gefunden werden. Das gestaltet sich meiner Erfahrung nach in ineffizienter Weise so, dass in Sprechstunden gemeinsame Überlegungen zur Datenbasis angestellt und das Vorgehen für jedes spezifische Projekt einzeln abgesprochen werden. Hierbei spielen im Wesentlichen die von Stukenbrock (2022, S. 318) zusammengefassten Fragen eine Rolle:

Welche Daten werde ich in welchem Format erheben? Welchen Umfang wird das Datenvolumen haben? Wie werde ich die Daten speichern, zusätzlich sichern, verwalten (Ordnerstrukturen, Benennungen, Verlinkungen, etc.)? Welche Metadaten benötige ich, wie erhebe und systematisiere ich sie? Wie behandle ich ethische und rechtliche Fragen? Benötige ich z.B. einen Passwortschutz für den Speicherort der Daten? Wie verfare ich mit dem Problem der langfristigen Aufbewahrung der Daten?

Die Antworten auf diese Fragen fallen dann aber zumeist approximativ aus, was Studierende durchaus etwas ratlos zurücklässt.

Die entlang der oben genannten Fragen entwickelte INSTAB-Formel ist ein Vorschlag für ein sachbezogenes, vereinheitlichendes Vorgehen, das Studierende in der Phase, in der noch keine Standards entwickelt worden sind, beim Verfolgen ihrer Forschungsinteressen unterstützen soll. Dabei steht *IN* für *In Time*, also den frühzeitigen Beginn der Datensammlung (2.1), *S* für *Speichern* (2.2), *T* für *Transkribieren* (2.3), *A* für *Annotieren* (2.4) und *B* für *Bereitstellen* (2.5). Studierende und ggf. auch Dozierende erhalten damit Anhaltspunkte für die Generierung von Datensammlungen im Rahmen von akademischen Lehrveranstaltungen, die soziopragmatische Phänomene zum Thema haben und auf hermeneutische Verfahren angewiesen sind. Ein besonderes Augenmerk liegt dabei darauf, die Kommunikate möglichst in ihrer Multimodalität abzubilden, weil die Selektion einzelner Modi der Komplexität der Daten in ihrer Medialität nicht gerecht wird. Dabei wurde berücksichtigt, dass der Prozess in einem überschaubaren Zeitraum mit zumutbarer Workload umsetzbar sein muss. Für das vorgeschlagene Verfahren werden keinerlei Programmierkenntnisse vorausgesetzt. Ich stelle hier eine ausschließlich auf On-Screen-Methoden (dazu Pfurtscheller 2022, S. 192) basierende Vorgehensweise vor. Das ist vermutlich nicht der eleganteste Weg, aber er hat den Vorteil, auf dem kleinsten gemeinsamen Nenner hinsichtlich der technischen Voraussetzungen, als auch der Computerkenntnisse aufbauen zu können.

2. Die INSTAB-Formel

2.1 IN wie In Time: Frühzeitig beginnen

Die Erstellung einer Datengrundlage ist zeitaufwendig und sollte daher bereits in der Vorlesungszeit (und bei Abschlussarbeiten vor der Anmeldung beim Prüfungsamt) begonnen werden. Für die Menge der Daten haben wir in Marx/Weidacher (2020, S. 33) die Faustformel formuliert, „dass genügend Daten vorhanden sind, so-

bald sich ein Muster finden lässt⁶. Es hat sich jedoch gezeigt, dass Studierende dennoch Richtwerte benötigen und eine Größenordnung von 50 bis 100 Belegen praxistauglich ist. Zu berücksichtigen ist hierbei, dass die für eine methodisch valide Untersuchung notwendige Stichprobengröße von diversen statistischen Parametern, wie z.B. erwartbaren Wahrscheinlichkeiten, dem Konfidenzlevel etc. abhängt und die genannten Richtwerte eher als unterste Grenze für empirische Einstiegsübungen im Rahmen studentischer Projekte dienen. Die Auswahl der Instagram-Profile hängt von der Forschungsfrage ab, die ebenfalls frühzeitig im Rahmen der Lehrveranstaltung festgelegt werden kann. Ein vollständiges Datum setzt sich aus den folgenden Komponenten zusammen:

- Angaben zum Profil (Profilfoto, Nutzername, Profilname, Bio (Kurzangaben zum* zur Profilinhaber*in, inkl. ggf. URL, Anzahl der Beiträge, Follower*innen und gefolgt Personen),
- Beitrag bestehend aus Foto, Caption inklusive Hashtags, Reaktionen und Kommentaren, ggf. Story oder Reel und
- Datum und ggf. Uhrzeit der Veröffentlichung.

Gut umsetzbar ist die Erhebung, wenn ein Protokollierungszeitraum von zwei Monaten festgelegt wird und die Daten jeweils täglich zu zwei festgelegten Uhrzeiten erhoben werden. Von jedem Beitrag sollte ein Screenshot und mit Hilfe der Instagram-App-Funktion ein QR-Code erstellt sowie Textinhalte per Copy-and-Paste extrahiert werden. Bilder, Reels und Stories können entweder über die App (z. B. unter Zuhilfenahme des Kurzbefehls R↓Download⁶ oder der App Instdown) oder browserbasiert⁷ heruntergeladen werden.

2.2 S wie Speichern

Ähnlich wie beim Crawlen mit dem Instaloader liegen die Kommunikat-Komponenten separat vor. Der automatisch im Ordner *Fotos* abgelegte QR-Code sollte wie die Bilder, Reels und Stories in der jeweiligen Uni-Cloud (z. B. Nextcloud) gespeichert werden. Alle Bild- und Videodateien werden dafür mit möglichst aussagekräftigen Dateinamen versehen, etwa nach dem Muster

- (1) Laufende Belegnummer_Format_Profil_Datum_Aufnahmenummer

Die laufende Belegnummer wird vergeben, damit alle Kommunikat-Komponenten problemlos aufeinander bezogen werden können. Die Angabe zum Format macht

⁶ www.youtube.com/watch?v=t5_CfErp-u8 (Stand: 2.4.2023).

⁷ Das lässt sich zum Beispiel für Bilder, Stories und Reels auf www.save-insta.com/de/ (Stand: 2.4.2023) sehr intuitiv bewerkstelligen. Besonders praktisch ist, dass z. B. die einzelnen Sequenzen von Stories separat herunterladbar sind.




Protokolltag	Belegnr.	Sigle	V-Datum Uhrzeit	Aufnahme QR-Code	Profilname	Screenshot	Story	Reel	Bild	
11.09.22	1_1_PEP_3822	n.b.	03.08.22		pigesplatin g B: 167 F: 2.521 A: 715		/	/		
<p>Caption</p> <p>Du kennst Menschen, die das kennen, zum Beispiel dich selbst. Schick dir eine Postkarte. Und anderen, die du damit ebenfalls einen Moment lang aufhebern kannst.</p> <p>Zehn Karten für 15 Euro bei frohmannweg.de</p> <p>Bildbeschreibung: Detail aus präraffaelischem Gemälde: Menschliche Figur mit langen lockigen Haaren in historischem orangem Gewand, blick streng. Hintergrund: Text: Den Text zu Fazit schreiben</p>										
Hashtags						Likes	Komm.-Nr.	Kommentar*innen	Kommentare	Likes
						117				
						#schreiben #buchschreiben #autor #autor:in #autor_in #menschen #bookstagram #community #postkarten #instagram	1	blanca.koerner:		Kenn ich, mach ich. 2
							2	Rhymesandit22	Awesome! 1	0

Abb. 1: Muster für die Dokumentation der Datenbasis, aus satztechnischen Gründen hier zweigeteilt dargestellt

schnell sichtbar, ob es sich um den QR-Code, eine Story, ein Reel oder ein Beitragsbild handelt. Mit Datum und (wenn verfügbar) Uhrzeit werden Angaben zum Veröffentlichungszeitpunkt des Posts vermerkt. Auch das Datum des Protokolltags ist festzuhalten. Der Profilname sollte deshalb mit in den Dateinamen aufgenommen werden, weil eine schnelle Zuweisbarkeit die Analysearbeit erleichtert. Hier sollte zusätzlich der Aktivitäts- und Vernetzungsgrad notiert werden, wofür die Anzahl der Follower (F), der abonnierten Accounts (A) und der Beiträge (B) Indikatoren sind. Die Aufnahmeummer gibt an, ob es sich bei dem Beleg um den ersten oder zweiten an dem angegebenen Tag protokollierten Beleg handelt. In ähnlicher Weise ist auch die Excel-Tabelle aufgebaut, in der alle Kommunikat-Komponenten zusammengeführt werden (siehe Abb. 1). Hierin werden auch Caption-Text, Hashtags und Kommentare übertragen. Für ephemere Inhalte, wie Stories und Reels, wird der Link integriert, der durch das Speichern in der Cloud erzeugt worden ist. Zudem ist es ratsam, eine Sigle zu erzeugen, um ggf. auch anonymisiert auf Belege aus der Datensammlung referieren zu können. Eine Möglichkeit zur Erstellung besteht beispielsweise darin, die laufende Belegnummer mit dem 1., 3. und 5. Buchstaben des Profilnamens sowie den Ziffern des Veröffentlichungsdatums zu kombinieren (siehe Abb. 1, Spalte 2).

Die Ablage in der Tabelle hat nicht nur den Vorteil, dass hier alle notwendigen Informationen zu einem Kommunikat zusammengeführt werden. Sie ist zudem durchsuchbar und beliebig erweiterbar. Das hat Soeffner (1989, S. 58) zufolge eine besondere Relevanz für die Annotation. Er betrachtet die Fixierung als Voraussetzung dafür, „dass etwas [überhaupt] zum ‚Datum‘ sozialwissenschaftlicher Analyse werden kann [und] immer wieder und in gleicher ‚Gestalt‘ von jedem beliebigen Interpreten angesehen, hin- und hergewendet und damit kontrolliert interpretiert werden kann“. Damit ist das Speichern der Daten nicht nur Grundlage für die Interpretation, sondern auch eine „Möglichkeit der Qualitätskontrolle qualitativer Forschung“ (Reichertz 2014, S. 66).

2.3 T wie Transkribieren

Während Textpassagen, wie Captions, Kommentare, Hashtags aber auch Texte in Stories oder Reels, recht einfach (per copy + paste-Verfahren) übernommen werden können, bestehen z. B. Stories und Reels aus bildlichen und akustischen (gesprochenen/gesungenen) Daten, die einer Transkription bedürfen. Damit stehen Studierende vor der Aufgabe der multimodalen medienlinguistischen Transkription, für die es nach wie vor keine einheitlichen Standards gibt (Pfurtscheller 2022, S. 187; siehe aber auch Stukenbrock 2009, S. 146; Reichertz 2014, S. 68). Stukenbrock (2009, S. 147f.) nimmt die digitale Verfügbarkeit und Komplexität der Daten zum Anlass darüber nachzudenken, dass ein Transkript allein keine adäquate Abbildung mehr darstellen kann. Das macht eine didaktische Reduktion für diesen wichtigen Brückenschlag zur Analysearbeit zur Herausforderung. Es geht in diesem Schritt um

nicht weniger als die Überführung (ephemer)-multimodaler Daten in ein Textformat. Um Studierende gedanklich zu entlasten, sei erwähnt, dass in der Forschung Einigkeit darüber besteht, dass (Bewegt-)Bildinhalte nicht verlustfrei in Text „übersetzt“ werden können, weshalb Reichertz (2014, S. 61–64) zum Beispiel vorschlägt, sich bei der Erzeugung einer Repräsentation von Medieninhalten abzuwenden und eher nach der Möglichkeit zu fragen, wie „die Bedeutung der Medieninhalte mit Hilfe eines anderen Mediums“ repräsentiert werden kann. Pfurtscheller (2022, S. 187 f.) rückt zentrale medienlinguistische Fragen in den Mittelpunkt seiner Überlegungen zu Transkriptions- und Annotationspraktiken in der qualitativen Forschung zu digitaler Medienkommunikation: Wie lässt sich Sprache in den Medien als mediatisierter Sprachgebrauch beschreiben? Welche Rolle spielt der Bestand von unterschiedlichen Kommunikations- und Gestaltungsmitteln, wie lassen sich Verfahrensweisen in Mikroanalysen rekonstruieren, welche Muster lassen sich in Mediendiskursen und im medialen Wandel beschreiben? Diese Fragen zeigen nicht nur eine Linie auf, entlang derer „kleinere“ für Seminar- und Abschlussarbeiten adäquate Fragestellungen entwickelt werden können, sie bilden auch das Grundgerüst für die Transkription und Annotation, bei der als wichtigstes Prinzip lediglich berücksichtigt wird, was für die individuelle Fragestellung relevant ist.

Für die hier zusammengetragenen Hinweise musste von diesen potenziellen Fragestellungen abstrahiert werden. Ich zeige daher ein etwas breiteres Spektrum an Transkriptionsoptionen auf, aus dem für die eigene Studie dann jeweils das Passende ausgewählt werden kann. Es sind praktische Gründe, die mich dazu veranlassen, die Arbeitsschritte Transkription und Annotation (siehe 2.4) voneinander zu trennen. Dabei orientiere ich mich an der von Pfurtscheller (2022, S. 189) explizierten Unterscheidung, nach der Transkription als Umwandeln und regelgeleitetes Inform-Bringen gefasst wird, während bei der Annotation „deskriptive oder analytische Metadaten oder Kategorien in rohes Datenmaterial“ eingebunden werden. Auf diese Weise können bereits im Datenaufbereitungsprozess Modi analytisch separiert werden. Ich schlage also vor, sich das in der Konversationsanalyse bewährte Zwiebelprinzip (Selting et al. 2009, S. 356) zunutze zu machen. In einem ersten Schritt werden den Videos allgemeine Kategorien zugewiesen, dann Bildsequenzen festgelegt (Schritt 2) und diese nach Bildinhalt (Schritt 3) aufgeschlüsselt sowie in Schritt 4 ggf. und selektiv multimodal erweiterte GAT2-Transkripte angefertigt. Was bedeutet das konkret?

Schritt 1: In Unkenntnis der jeweils für die Studienarbeit zusammengestellten Datenbasis greife ich für Vorschläge zur Unterteilung der Instagram-Videos auf Bainotti/Caliandro/Gandini (2021, S. 3664) zurück, die folgende beschreibende Kategorien ermittelt haben: Porträts (von sich selbst, den Freunden, der Familie etc.), Kompositionen (Texte, Bilder, Sticker, Memes, Zeichnungen etc.), Objekte (Autos, Bücher, Körper etc.), Szenerien (Landschaft/Natur, Städte, Reisen etc.), Feierlich-

keiten (Party, Geburtstag, Feiertage etc.), Ernährung (Essen, Getränke), Tiere und anderes.

Schritt 2: Sequenzen, die durch die jeweilige Story (Collage) oder Szenen in Reels vorgegeben werden, können als grobe Strukturierung übernommen werden. Unterschiedliche Bilder in der Story werden also ebenso wie unterschiedliche Szenen in Reels getrennt voneinander transkribiert und jeweils fortlaufend nummeriert.

Schritt 3: Anschließend werden die Inhalte der Sequenzen und Szenen sprachlich paraphrasiert. Dazu werden die folgenden Fragen beantwortet: Wer oder was ist zu sehen? Was tut ggf. die Person/was tun die Personen? Welche akustischen Signale gibt es, z. B. Monologe, Gespräche oder Musik? Wurden Filter, Rahmen, Sticker eingesetzt? Wird auf Links verwiesen, gibt es sogenannte Shoutouts, also Erwähnungen anderer Nutzer*innen/Follower*innen?

Schritt 4: Für die Fragestellung relevante Monologe und/oder Gespräche werden in ein multimodal erweitertes GAT2-Transkript übertragen. Hierbei dienen Sprechen und Pausen als Gerüst und werden fettgedruckt. Körperliche Aktivitäten, Geräusche, visuelle und textuelle Effekte werden mit Sonderzeichen aligniert, vgl. dazu Mondada (2016), sowie Marx/Schmidt (2019, S. 329). In Abbildung 2 habe ich das einmal für die zweite Szene einer Story umgesetzt. Diese Sequenz wäre zum Beispiel interessant, wenn man sogenannte Call-to-actions-Handlungen untersuchen möchte, die typisch für Instagram sind.

INS	+unten %~und \$zwar (.)(*h) §DEN hier von Veganista;
kA ⁸	% greift nach Gutschein § schaut auf Gutschein § zeigt Gutschein und blickt in Kamera, beugt sich nach vorn --->>
StA	+ Videobeschreibung --->>
HG ⁹	~ rascheln
INS	mit dem man ~sich eine % \$ein Liter EIS §box holen kann(*h);
kA	% schaut kurz auf Gutschein § zeigt Gutschein und blickt in Kamera § fächernde Bewegung mit Gutschein
StA	--->>
HG	~ leichtes rascheln
INS	%und ich HAB mir gedacht –
kA	% dreht Gutschein und schaut auf diesen
StA	--->>
HG	

⁸ körperliche Aktivität

⁹ Hintergrundgeräusche

350 Konstanze Marx

INS %ich verlos \$~ den hier in der Story an jemanden §unter EUCH –
kA % blickt in Kamera
§ klatscht einmal in Hände

§ schaut auf Gutschein

StA --->>

HG ~ Klatschgeräusch

INS %also (*h) viel \$~leicht mach ich damit ja jemandem eine FREUde –

kA % blickt in Kamera
§ klatscht einmal in Hände

StA --->>

HG ~ Klatschgeräusch

INS alle Bedingungen gibt es in der nächsten %SLIDE;

kA % „wischt“ rechte Hand von oben rechts
nach unten links

StA --->>

HG

INS und ICH geh jetzt % raus.

kA % greift mit linker Hand nach vorn

StA --->>

HG

Abb. 2: Muster für ein multimodal erweitertes Transkript



Sind die Daten letztlich für die Analyse aufbereitenden Schritte vollzogen, kann zur Annotation übergegangen werden. Es bietet sich an, die unter Punkt 3 und Punkt 4 beschriebenen Vorgänge in separaten Spalten in die bereits erstellte Datenbasis (siehe Abb. 1) zu integrieren.

2.4 A wie Annotieren

Die Annotation dient dazu, für die Analyse wichtige Beobachtungen zu notieren. Ganz einfach formuliert, wird hier festgehalten, was auffällt. Das kann auf unterschiedlichen Ebenen geschehen. Ich nenne einige Beispiele.

In den textbasierten Passagen sind Auffälligkeiten auf allen linguistischen Beschreibungsebenen interessant: Gibt es besondere Wörter, auch mit Blick auf Hashtagkonnektive,¹⁰ fällt der Satzbau auf, wie sind die Sätze und Hashtags miteinander verknüpft? Wurden Emojis verwendet, welche und wo genau?

¹⁰ Hashtagkonnektive sind neben dem Rautenmarker konstitutive Bestandteile von Hashtags. Das können vielgestaltige Formen sein, etwa Akronyme, Zahlen, einzelne oder mehrere Wörter oder sogar Sätze (Marx 2021, S. 135).

Bei Bildern helfen Kameraperspektive, Licht, eingesetzte Filter oder Effekte, die bildliche Ausdruckshandlung zu ergründen. Reichertz (2014, S. 69) schlägt vor, mit sogenannten *moves* als kleinste bedeutungstragende Bewegungen zu arbeiten, um weg vom Bild, das im Transkriptionsschritt beschrieben wurde, auf die Deutungsebene zu gelangen. Für Instagram-Videos legen Bainotti/Caliandro/Gandini (2021, S. 3654) Narrationstypen auf einem sogenannten „connotative level“ fest, weil sie davon ausgehen, dass es sich hierbei um digitale *small stories* handelt. Small stories bieten die Möglichkeit, alltägliche, gewöhnliche, sogar triviale Ereignisse (Georgakopoulou 2017, S. 268) und spezifische kulturelle und moralische Sichtweisen, anders als in traditionellen Geschichten, „a-typisch“, d. h. fragmentarisch, mit offenem Ende und mit hoher intertextueller Verankerung zu erzählen (Page 2013; Bainotti/Caliandro/Gandini 2021, S. 3654). Als Kategorien wurden hierbei ermittelt: Alltagseinblicke in einzelnen Einheiten (sog. *snippets*) oder kohärente Erzählungen eines Ereignisses in mehreren Szenen/Frames als Mikrodokumentationen. Vorgeschlagen wird, den jeweiligen Verwendungszusammenhang (besonderes Ereignis, Alltag, Stimmung, Interaktion, Selbstdarstellung, Aphorismus o. ä.) zu notieren. Bei Interaktant*innen (etwa Personen oder Tiere) sollte festgehalten werden, wie sich diese Interaktion gestaltet, ob es Auffälligkeiten mit Blick auf die Darstellung ihrer Beziehung, der verhandelten Themen etc. gibt. Werden z. B. Emotionen ausgedrückt, wenn ja, welche und wie? Interessant kann auch sein, ob es besondere Schnitte oder ein spezifisches Tempo gibt und welche Stimmung etwaige Musik konstruiert. Gibt es besondere Symbole, eine auffällige Wort-/Farb-/Typografie-Wahl bei Texteinblendungen? Inwieweit nimmt auch der*die Profilinghaber*in dazu eine Position ein und wie? Dabei kann es u. a. eine Rolle spielen, ob die jeweilige Story zu den Highlights hinzugefügt wurde.

Die Kodierung sollte hier so erfolgen, dass sie von Dritten leicht nachvollzogen werden kann. Ich rate also davon ab, umständliche, auf vorab definierten Ikons basierende Notationssysteme zu verwenden (siehe auch Reichertz 2014, S. 69). Dass Annotationen im Normalfall von mehr als einer Person durchgeführt werden, sollte Studierenden bewusst sein. Dieses methodische Wissen kann im Methodenkapitel erwähnt werden, eine Umsetzung dieser Praxis liegt nicht im Erwartungshorizont für studentische Arbeiten.

2.5 B wie Bereitstellen

Wie soll das Datenmaterial für Dozierende zugänglich gemacht werden? Fragen, die hierzu von Studierenden häufig gestellt werden, beziehen sich a) auf die konkrete Umsetzung des Zugangs zu den Daten und b) auf ggf. zu berücksichtigende ethische Aspekte.

In vielen Prüfungsordnungen ist nach wie vor vorgeschrieben, dass Seminar- und Abschlussarbeiten als ausgedrucktes Manuskript einzureichen sind. In manchen

Universitäten wird zusätzlich eine CD-ROM eingefordert. Solche Vorgaben bedürfen einer dringenden Überarbeitung. Solange sie jedoch verbindlich sind, empfehle ich zusätzlich eine lesefreundliche Variante der Bereitstellung. Das heißt, dass die Daten (als illustrierende Beispiele etwa) sowohl in den Fließtext, als auch in den Anhang so eingebunden sein sollten, dass ein schneller Zugriff erfolgen kann. Für textuelle Daten und Screenshots lässt sich das leicht bewerkstelligen. Aber auch ephemere Daten können eingebunden und leicht zugänglich gemacht werden. Im Abschnitt 2.2 habe ich bereits das Erstellen von QR-Codes über die Instagram-App empfohlen. Diese lassen sich problemlos in den Fließtext integrieren. Für die Bereitstellung der gesamten Datengrundlage erachte ich es als praktikabel, Zugang zu dem Ordner zu gewähren, der in der universitätseigenen Cloud angelegt ist (siehe 2.2). Das kann über einen Link oder ebenfalls über einen QR-Code¹¹ geschehen. Die Ordnerstruktur sollte dabei leicht nachvollziehbar gestaltet, die einzelnen Unterordner und Dateien sollten mit transparenten Namen (etwa Name_Modul_Angabedes-Semesters_Datengrundlage_gesamt, _Stories, _Reels etc.) versehen sein.

Die hier zusammengetragenen Hinweise sind an qualitativen Untersuchungen ausgerichtet, die häufig Orientierungspunkt für forschungsethische Fragestellungen sind (vgl. Meier-Vieracker 2022, S. 20). Daher sollte auch in Seminar- und Abschlussarbeiten sichtbar gemacht werden, dass sich die Verfasser*innen mit ethischen Fragen auseinandergesetzt haben. Eine solche Passage kann z. B. gut im Methoden-Kapitel platziert werden und sollte die wesentlichen Überlegungen zur Anonymisierung, zur (informationellen) Selbstbestimmung mit Blick auf die Urheber*innen und zur Datentransparenz enthalten. Die Daten (insbesondere bei Seminararbeiten) werden im Normalfall von zwei Personen, nämlich den Betreuer*innen der Arbeit, eingehend gelesen. Diese Tatsache sollte in die Überlegung, ob umfangreiche Anonymisierungen innerhalb der zugänglich gemachten Datenbasis notwendig sind, einfließen und kann in der erwähnten Passage explizit thematisiert werden. Vor diesem Hintergrund und unter Berücksichtigung des Umstands, dass Studierenden nur ein sehr begrenzter Zeitraum für das Erstellen von Seminararbeiten (aber natürlich auch Abschlussarbeiten) zur Verfügung steht, ist auch abzuwägen, ob auf die sehr aufwändige, für wissenschaftliche Projekte empfohlene Kontaktaufnahme zu Nutzer*innen (vgl. Luth/Marx/Pentzold 2022, S. 120) verzichtet werden kann. Das betrifft auch die in diesem Zusammenhang als Alternative angeführte Datenakquise über Spenden. Es ist nicht davon auszugehen, dass solche forschungsethisch vorzuziehenden Varianten im Rahmen von studentischen Arbeiten umsetzbar sind. Das Bewusstsein für diese Problematik sollte aber kenntlich gemacht werden. Eine Formulierung könnte folgendermaßen aussehen:

Die dieser Arbeit zugrundeliegenden Daten sind einem Sozialen Netzwerk entnommen, das ohne besondere Hürden öffentlich zugänglich ist. Mir ist bewusst, dass dar-

¹¹ Dieser lässt sich z. B. mit dem QR-Code-Generator leicht erstellen: www.qrcode-generator.de/ (Stand: 2.4.2023).

aus noch nicht abzuleiten ist, dass die Nutzer*innen einer wissenschaftlichen Untersuchung ihrer Beiträge zugestimmt haben. Ich gehe davon aus, dass diese Arbeit nur von meinen Betreuer*innen gelesen wird, auch die Datengrundlage wird nur ihnen zugänglich gemacht. Ich habe mit Blick auf den mir zur Verfügung stehenden zeitlichen Rahmen zur Anfertigung meiner Seminararbeit auf umfangreiche Anonymisierungen verzichtet und von einer Kontaktaufnahme zu den betreffenden Nutzer*innen abgesehen. Auch die Akquise der Daten über Spenden wäre in diesem Zusammenhang nicht umsetzbar gewesen. Sollte meine Untersuchung qualitativ so überzeugen, dass eine Veröffentlichung in Betracht kommt, werde ich forschungsethische Aspekte erneut in den Blick nehmen.

3. Fazit

Ziel dieses Beitrags war eine Gebrauchsanweisung für Datensammlungen, die Basis für qualitativ ausgerichtete studentische Arbeiten sein kann. Damit schließt er in bescheidener Weise an die umfassenden Arbeiten von Angelika Storrer zur Generierung von Social-Media-Korpora an. Ich habe hier versucht, ein Desiderat zu adressieren, das die Gemüter in großen und bereits lange währenden Forschungsdiskussionen bewegt und uns im Kleinen in unserer täglichen Arbeit, der akademischen Lehre, vor ganz konkrete Herausforderungen stellt: Die Frage, wie multimodale Daten als Grundlage für studentische Forschungsprojekte abgebildet werden können. Zu berücksichtigen ist hierbei, dass deren Generierung fragengeleitet und in einer angemessenen Zeit umsetzbar sein muss, möglichst keine Programmierkenntnisse voraussetzen und nur von einem technischen Grundverständnis bei der Handhabung verfügbarer Software ausgehen darf. Vielmehr können diejenigen Fähigkeiten zum Einsatz kommen, die ohnehin zur Nutzung von Social Media benötigt werden. Mit der INSTAB-Formel wurde ein didaktisch reduzierter Vorschlag vorgelegt, der meines Erachtens auch deshalb legitim ist, weil studentische Arbeiten häufig (und im wertschätzenden Sinne: nur) hypothesengenerierenden Charakter für Qualifikationsarbeiten haben. Wünschenswert ist natürlich, dass der Anwendungsbezug dazu beiträgt, dass der Vorschlag über den festlichen Rahmen dieses Buches hinaus Reichweite erlangt, weil das der Würdigung des Werks von Angelika Storrer umso zuträglicher wäre.

Literatur

- Achmann, Michael/Hampel, Lisa/Asabidi, Ruslan/Wolff, Christian (2022): Studying the ephemeral cultures of digital oblivion. Identifying patterns in Instagram Stories. In: Geierhos/Trilcke/Börner/Seifert/Busch/Helling (Hg.).
- Bainotti, Lucia/Caliandro, Alessandro/Gandini, Alessandro (2021): From archive cultures to ephemeral content, and back: studying Instagram stories with digital methods. In: *New Media & Society* 23, 12, S. 3656–3676.
- Bateman, John A. (2008): *Multimodality and Genre. A Foundation for the Systematic Analysis of Multimodal Documents*. New York: Palgrave Macmillan.

- Beißwenger, Michael/Lüngen, Harald (2022): Korpora internetbasierter Kommunikation. In: Beißwenger/Lemnitzer/Müller-Spitzer (Hg.), S. 431–448.
- Beißwenger, Michael/Lemnitzer, Lothar/Müller-Spitzer, Carolin (Hg.) (2022): Forschen in der Linguistik. Eine Methodeneinführung für das Germanistik-Studium. (= UTB 5711). Paderborn: Brill|Fink.
- Geierhos, Michaela/Trilcke, Peer/Börner, Ingo/Seifert, Sabine/Busch, Anna/Helling, Patrick (Hg.) (2022): Eine Posterpräsentation auf der 8. Tagung des Verbands „Digital Humanities im deutschsprachigen Raum“ – DHd 2022 Kulturen des digitalen Gedächtnisses, Potsdam, 07.03.–11.03.2022.
- Georgakopoulou, Alexandra (2017): Small stories research: a narrative paradigm for the analysis of social media. In: Sloan, Luke/Quan-Haase, Anabel (Hg.): The SAGE handbook of social media research methods. London u. a.: SAGE, S. 266–281.
- Gredel, Eva (Hg.) (2022): Diskurse – digital. Theorien, Methoden, Anwendungen. (= Diskursmuster – Discourse Patterns 30). Berlin/Boston: De Gruyter.
- Gür-Şeker, Derya (2021a): Zur Konstruktion rechter Diskursgemeinschaften in Social Media. Von PEGIDA zur Identitären Bewegung. In: Lublin Studies in modern languages and literature 45, 2, S. 119–131.
- Gür-Şeker, Derya (2021b): „Wie sieht die #ZukunftDerArbeit aus?“ – Hashtags als Dreh- und Angelpunkte linguistischer Social-Media-Analysen. In: *Linguistica* 61, 1, S. 67–80.
- Heyen, Hauke (2022): #hokerbeest: Auf der Suche nach Spuren digitaler nordfriesischer Kommunikation. In: Walker, Alastair/Hoekstra, Eric/Jensma, Goffe/Vanselow, Wendy/Visser, Willem/Winter, Christoph (Hg.): From West to North Frisia. A journey along the North Sea Coast. Frisian studies in honour of Jarich Hoekstra. (= NOWELE Supplement Series 33). Amsterdam/Philadelphia: Benjamins, S. 133–148.
- Issel-Dombert, Sandra (2020): Die Macht der Bilder und Bilder der Macht: die sprachliche Konstruktion von *Nationalismus* und *Konservatismus* in der strategischen politischen Kommunikation von *Vox España*. In: Zeitschrift für romanische Philologie 136, 2, S. 538–566.
- Knierim, Aenne/Achmann, Michael/Wolff, Christian (2022): Zeitgeschichte untersuchen – Topic Modeling von #blackouttuesday-Inhalten auf Instagram. In: Geierhos/Trilcke/Börner/Seifert/Busch/Helling (Hg.).
- Krammer, Lisa (2021): „Wir sind für euch da. #gemeinsamschaffenwirdas“ Flugblätter-Narrative und Social-Media-Kommunikation des Lebensmittelhandels während des ersten COVID-19-Lockdowns in Österreich. In: Wiener Linguistische Gazette (WLG) 90, 2021 (Themenheft: Corona-Diskurse in und über Österreich), S. 185–216.
- Luth, Janine/Marx, Konstanze/Pentzold, Christian (2022): Ethische und rechtliche Aspekte der Analyse von digitalen Diskursen. In: Gredel (Hg.), S. 101–134.
- Marx, Konstanze (2021): #Versprochenistversprochen – Die Kartographie einer kognitionslinguistischen Rautenwanderung. In: Giessen, Hans W./Lüger, Heinz-Helmut (Hg.): Text-, Diskurs- und Kommunikationsforschung. Festschrift für Hartmut Lenk. (= Landauer Schriften zur Kommunikations- und Kulturwissenschaft). Landau: Verlag Empirische Pädagogik, S. 131–144.
- Marx, Konstanze/Schmidt, Axel (2019): Making Let's Plays watchable: Praktiken des stellvertretenden Erlebbar-Machens von Interaktivität in vorgeführten Videospielen. In: Marx,

- Konstanze/Schmidt, Axel (Hg.): *Interaktion und Medien.* (= OraLingua 17). Heidelberg: Winter, S. 319–352.
- Marx, Konstanze/Weidacher, Georg (2020): *Internetlinguistik: ein Lehr- und Arbeitsbuch.* 2., aktual. und durchges. Ausgabe. (= Narr Studienbücher). Tübingen: Narr.
- Meier-Vieracker, Simon (2022): *Digitale Diskursforschung und Forschungen zu digitalen Diskursen.* In: Gredel (Hg.), S. 7–26.
- Mondada, Lorenza (2016): *Zwischen Text und Bild: Multimodale Transkription.* In: Hausendorf, Heiko/Schmitt, Reinhold/Kesselheim, Wolfgang (Hg.): *Interaktionsarchitektur, Sozialtopographie und Interaktionsraum.* (= Studien zur Deutschen Sprache 72). Tübingen: Narr, S. 111–160.
- Page, Ruth E. (2013): *Stories and Social Media: identities and interaction.* (= Routledge studies in sociolinguistics 3). New York u. a.: Routledge.
- Pfurtscheller, Daniel (2022): *Medienlinguistisch transkribieren. Transkriptions- und Annotationspraktiken in der qualitativen Forschung zu digitaler Medienkommunikation.* In: Schwarze, Cordula/Grawunder, Sven (Hg.): *Transkription und Annotation gesprochener Sprache und multimodaler Interaktion. Konzepte, Probleme, Lösungen.* Tübingen: Narr, S. 186–210.
- Reichertz, Jo (2014): *Das vertextete Bild. Überlegungen zur Gültigkeit von Videoanalysen.* In: Moritz, Christine (Hg.): *Transkription von Video- und Filmdaten in der Qualitativen Sozialforschung. Multidisziplinäre Annäherungen an einen komplexen Datentypus.* Wiesbaden: Springer, S. 55–72.
- Schiefer, Veronika (2022): *Der Instagram-Beitrag als politische Textsorte. Eine Analyse rechtspopulistischer Merkmale in Instagram-Beiträgen der FPÖ im Kontext der Corona-Krise.* Masterarbeit. Graz: Universität Graz.
- Selting, Margret/Auer, Peter/Barth-Weingarten, Dagmar/Bergmann, Jörg/Bergmann, Pia/Birkner, Karin/Couper-Kuhlen, Elizabeth/Deppermann, Arnulf/Gilles, Peter/Günthner, Susanne/Hartung, Martin/Kern, Friederike/Mertzluff, Christine/Meyer, Christian/Morek, Miriam/Oberzaucher, Frank/Peters, Jörg/Quasthoff, Uta/Schütte, Wilfried/Stukenbrock, Anja/Uhmann, Susanne (2009): *Gesprächsanalytisches Transkriptionssystem 2 (GAT 2).* In: *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10, S. 353–402.
- Soeffner, Hans-Georg (1989): *Auslegung des Alltags – Der Alltag der Auslegung. Zur wissenssoziologischen Konzeption einer sozialwissenschaftlichen Hermeneutik.* Unter redaktioneller Mitarbeit von Ludgera Vogt. Frankfurt a. M.: Suhrkamp.
- Stukenbrock, Anja (2009): *Herausforderungen der multimodalen Transkription: Methodische und theoretische Überlegungen aus der wissenschaftlichen Praxis.* In: Birkner, Karin/Stukenbrock Anja (Hg.): *Die Arbeit mit Transkripten in Fortbildung, Lehre und Forschung.* Mannheim: Verlag für Gesprächsforschung, S. 144–169.
- Stukenbrock, Anja (2022): *Audio- und Videographie.* In: Beißwenger/Lemnitzer/Müller-Spitzer (Hg.), S. 307–323.
- Vargas, Francielle/Carvalho, Isabelle/Rodrigues de Góes, Fabiana/Benevenuto, Fabrício/Pardo, Thiago A. S. (2022): *HateBR: a large expert annotated corpus of Brazilian Instagram comments for hate speech and offensive language detection.* In: Calzolari, Nicoletta/Béchet, Frédéric/Blache, Philippe/Choukri, Khalid/Cieri, Christopher/Declerck, Thierry/Goggi, Sara/Isahara, Hitoshi/Maegaard, Bente/Mariani, Joseph/Mazo, Hélène/Odijk, Jan/

Piperidis, Stelios (Hg.): Proceedings of the thirteenth language resources and evaluation conference, Marseille, France. European Language Resources Association, S. 7174–7183. <https://arxiv.org/pdf/2103.14972.pdf> (Stand: 2.4.2023).

Weißer, Martin (2022): Python-Programmierung für Germanist:innen. Ein Lehr- und Arbeitsbuch. (= Narr Studienbücher). Tübingen: Narr.