

# **Digitale Narrative zu Wissenschaftler\*innen**

## Selbstdarstellung in sozialen Medien

---

*Johanna Hartmann, Mareike Schuhmacher<sup>1</sup>, Annalena Hiergeist, Judith Ackermann<sup>2</sup>*

**Abstract:** Die Studie analysiert 21.811 Twitter-Accounts von Wissenschaftler\*innen und ihre Tweets mit Blick auf die digitale Selbstdarstellung nach Gender. Mit digitalen Methoden (u. a. Stilometrie, Topic Modeling, Word Embeddings) wird herausgearbeitet, dass Wissenschaftler\*innen\* im Sinne des FLINTA\*-Begriffs häufiger Care- und Genderthemen betonen und sich in den Geistes- und Sozialwissenschaften verorten, während männlich gelesene Accounts größere Nähe zu Technikwissenschaften und Ökonomie aufweisen und deutlich häufiger den Begriff »Experte« zur Selbstbeschreibung nutzen. Gender prägt somit digitale Selbstanarrative.

## **1 Hintergrund**

Soziale Medien erlauben Wissenschaftler\*innen, sich und ihre Forschung zu präsentieren (Côté und Darling 2018). Dies bietet Chancen für die eigene Karriere, etwa durch Teilhabe an digitalen Diskursen des Fachs, Informationsgewinn oder Vernetzungsmöglichkeiten (Cheplygina et al. 2020; Collins et al. 2016; Mohammadi et al. 2018). Insbesondere weibliche Early-Career-Forschende fungieren als nahbare Vorbilder und wirken Genderstereotypen entgegen (Phillips et al. 2022). Eine Social-Media-Präsenz kann sich zudem positiv auf die Zitation von Publikationen auswirken (Eysenbach 2011; Klar et al. 2020; Lamb et al. 2018).

Im Jahr 2020 war Twitter (im Juli 2023 umbenannt in X) im deutschsprachigen Wissenschaftsraum die wichtigste Plattform, um sich mit anderen Forschenden auszutauschen (23,5 %) und Informationen der Fachdisziplin zu gewinnen (45,4 %; Zahlen von Hennig und Kohler 2020). Studien, die Wissenschaftskommunikation mit Blick auf Gender analysieren, zeigen anhand von YouTube-Kommentaren, dass Frauen mehr kritische, sexistische oder feindliche Kommentare erhalten (Amarasekara und Grant 2019: 68). Zudem haben Kanäle mit weiblichen Moderatorinnen weniger Abonnements,

---

<sup>1</sup> Johanna Hartmann und Mareike Schuhmacher haben gleichwertig zum Manuscript beigetragen.  
<sup>2</sup> Judith Ackermann hat das Projekt und die Kooperation geleitet sowie ebenfalls wesentlich zur Konzeption, zur Analyse und zum Manuscript beigetragen.

erfahren aber ein höheres Engagement als die von Männern (ebd.). Allerdings könnten solche Befunde dadurch verzerrt werden, dass insgesamt weniger Frauen unter den Nutzenden sind und bei Befragungen Frauen seltener teilnehmen (vgl. Mohammadi et al. 2018).

Mit Blick auf die Rezeption von Tweets zeigt sich, dass Postings von Wissenschaftler\*innen auf kleinen Accounts zwar zunächst ein wissenschaftsnahes Publikum – oft sind darunter Forschende der eigenen Disziplin (Jünger und Fähnrich 2020) –, ab 1.000 Follower\*innen jedoch zunehmend auch Personen außerhalb der Wissenschaft erreichen (Côté und Darling 2018). Soziale Medien eignen sich demnach auch zur Kommunikation mit der breiteren Öffentlichkeit (ebd.). Das inhaltliche Spektrum geht mit mehr als einem Viertel Tweets zu politischen und sozialen Inhalten über die Forschung hinaus (ebd.) und die Accounts werden oft sowohl für private als auch für professionelle Zwecke verwendet (Bowman 2015).

Die Äußerungen der Wissenschaftler\*innen auf Social Media lassen sich zwischen »Professional Self Disclosure« (PFSD) und »Private Self Disclosure« (PSD) verorten (Zhang und Lu 2023). Während PFSD das Auftreten als Person im beruflichen Kontext und das Teilen wissenschaftlicher Inhalte meint, fokussiert PSD das Teilen freizeitlicher und nicht wissenschaftsbezogener Inhalte (ebd.). Derweil das Teilen persönlicher Informationen in Tweets keinen Einfluss auf die wahrgenommene Authentizität der Forschenden hat, steigert PFSD die Publikumsbindung und erlaubt den Aufbau stärkerer parasozialer Bindungen als bei Influencer\*innen aus anderen Themenbereichen (ebd.). Dabei wird Wissenschaftler\*innen auf Twitter im Bereich von Gesundheit und Umwelt mit ihrer Institutionszugehörigkeit vom Publikum eine größere Authentizität und mehr Expertise zugeschrieben als vergleichbaren Accounts von Influencer\*innen, die keine Wissenschaftler\*innen sind (ebd.; siehe auch Rohloff in diesem Band).

Die bisherige Forschung zum Thema (Selbst-)Darstellung von Wissenschaftler\*innen in den sozialen Medien zeigt ein facettenreiches Bild. Die große Mehrheit der genannten Studien beschäftigt sich allerdings mit beispielhaft ausgewählten kleinen und mittelgroßen Datensätzen und/oder nutzt Methoden der qualitativen Sozialforschung. Die vorliegende Studie fügt dem einen groß angelegten korpusbasierten Ansatz hinzu, bei dem digitale Methoden genutzt werden, um einen möglichst vollständigen Datensatz inhaltlich auszuwerten. Das Methodenspektrum reicht von der automatischen Klassifikation von Genderzuschreibung über die Untersuchung von Dependenz und die Analyse stilistischer Merkmale bis hin zu Machine-Learning-Methoden wie Topic Modeling und Word Embeddings. Im Zentrum steht die Forschungsfrage: Wie unterscheidet sich die (Selbst-)Darstellung von Forschenden auf Twitter nach Gender?

## Korpus

Im BMFTR-Projekt »Wissenschaftlerinnen in die Medien« wurden zwischen dem 14.12.2022 und dem 25.01.2023 Twitter-Accounts und Tweets zu den Hashtags »#Forschung« und »#PhDLife« gescraped<sup>3</sup>, um digital vermittelte Narrative über Wissen-

---

<sup>3</sup> Beim Scraping werden Informationen von einer Webseite über eine technische Schnittstelle ausgelesen und heruntergeladen, also sozusagen von der Webseite »gekratzt«.

schaftler\*innen zu erfassen. Zusätzlich wurde den Forschenden Zugriff auf die Tweets zum Hashtag »#IchBinHanna« gewährt, die im Zeitraum vom 10.06.2021 bis zum 13.04.2022 durch das #IchBinHanna Research Collective<sup>4</sup> gesammelt wurden.

Das daraus zusammengestellte Korpus umfasst 76.013 Accounts (davon 40.629 deutschsprachig) mit insgesamt 186.839 Tweets (davon 106.075 deutschsprachig). Neben Accounts von Personen sind auch solche von Institutionen und journalistischen Einrichtungen eingeschlossen. Aus der Gesamtmenge ließen sich 21.811 Accounts einzelner Wissenschaftler\*innen ermitteln, deren Gender teilautomatisiert mittels einer Mixed-Methods-Pipeline klassifiziert wurde. Auf diesen Accounts wurden die Biografien betrachtet, die meist eine kurze Selbstdarstellung enthalten. Die durchschnittliche Zeichenlänge dieser »Bios« beträgt 93,4 Zeichen, schwankt jedoch stark ( $sd = 58,9$ ). Eine zweite Teilmenge von 10.181 Accounts wurde manuell kategorisiert, um Inhalte (d. h. Tweets) von Wissenschaftler\*innen, wissenschaftsnahen Personen, Forschungsinstitutionen, Hochschulen und Universitäten vergleichend zu analysieren.

Wir analysieren sowohl englischsprachige als auch deutschsprachige Accounts, um einen Direktvergleich zu erreichen. Die Sprache der Account-Biografie sagt dabei nicht unbedingt etwas über den Wohn- oder Tätigkeitsort der Wissenschaftler\*innen aus – von den Accounts, die durch ihre Ortsangabe eindeutig Deutschland zugeordnet werden können, haben 60 % englischsprachige Biografien, 38 % deutsche und 2 % anderssprachige.

## 2 Methodisches Vorgehen

Digital-Humanities-Studien, die Gender klassifizieren und analysieren, arbeiten oft binär (z. B. Rybicki 2016; Freitas und Santos 2023; Vianne et al. 2023). Allerdings wird in einzelnen Studien deutlich, dass dies der Vielfalt von Gendern nicht gerecht wird. Underwood et al. fanden 2018 heraus, dass ab 1850 in englischsprachiger Literatur Männer- und Frauenfiguren immer weniger distinkтив beschrieben werden. Cheng (2020) betrachtet in einem ähnlichen Ansatz Gender als skalares Phänomen, das sich zwischen den Polen »männlich« und »weiblich« entfaltet. Anhand kleinerer Korpora wurde getestet, ob ein diverseres Genderkonzept als Mehrkategoriensystem für die digitale geisteswissenschaftliche Forschung fruchtbar gemacht werden kann (Schumacher und Flüh 2023). Auf dieser Basis wurde ein Machine-Learning-Classifier entwickelt (Schumacher 2021), der Figurenreferenzen in literarischen Texten den drei Kategorien »weiblich«, »männlich« und »neutral« zuordnen kann.

In der vorliegenden Studie wird Gender als sozio-kulturell konstruiert betrachtet (vgl. v. a. de Beauvoir 1949 und 1992; Butler 1996 und 2002; Connell 1996) und es wird von einer grundsätzlichen Diversität der Kategorien ausgegangen. Unter Anwendung

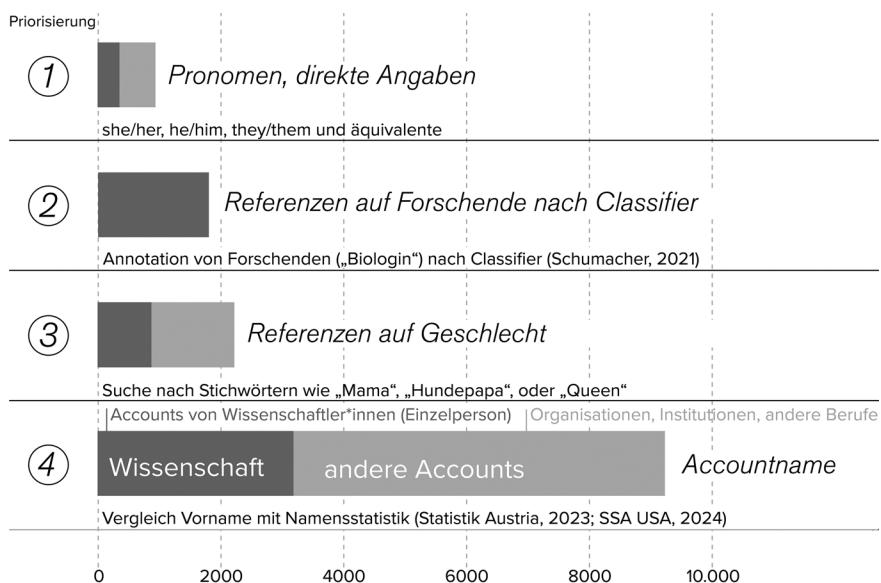
4 Das #IchBinHanna Research Collective ist »a collaborative, interdisciplinary endeavour of scholars from the German academic context. We look at Twitter data from the 2021 hashtag #IchbinHanna et al. and combine organising-specific concerns with academic social media research.« (Laser et al. 2021) Die Mitglieder des Kollektivs sind (in alphabetischer Reihenfolge): David Adler, Migle Barreikyte, Elen Le Foll, Christian Funk, Stefan Laser und Jana Lasser.

der Narrative-Identity-These von Ricoeur (1987) gehen wir davon aus, dass sich das Gender einer Person einerseits durch Ausdrücke zeigt, die diese verwendet, um sich selbst zu beschreiben, und andererseits durch Fremdzuschreibungen. Darum untersuchen wir zunächst Account-Biografien, die wir erst für eine Genderklassifikation der Accounts nutzen und in einem zweiten Schritt inhaltlich auswerten. Ergänzend werden die Tweets der ausgewählten und klassifizierten Accounts im Hinblick auf die Darstellung der Forschenden unterschiedlicher Gender betrachtet.

## 2.1 Genderklassifikation von Accounts

Den Ansätzen von Gender als sozialer Konstruktion und Teil einer narrativen Identität folgend, sind Pronomen und Nomen (inkl. Namen), die sich auf Personen beziehen, relevant für die Genderzuordnung, da sie meist Informationen zum Gender der bezeichneten Person enthalten. Darum erfolgte die Genderzuordnung der Accounts anhand von drei Indikatoren: Pronomen, nominale Referenzen – einmal mittels Gender-Classifier automatisch annotiert und einmal durch teilautomatisierte Wortsuche ergänzt – sowie Namen, die mit aktuellen Statistiken abgeglichen wurden (Abb. 1).

Abb. 1: Indikatoren für Genderzuordnungen nach Priorität und Häufigkeit



Zusätzlich wurden aus den klassifizierten Accounts durch Abfrage geeigneter Stichworte (z. B. »Universität«, »Postdoc«) solche herausgefiltert, die von Wissenschaftler\*innen und nicht von Institutionen geführt werden. Insgesamt konnten 4.020 Accounts von Wissenschaftler\*innen mit Bios in deutscher und 17.791 Accounts in englischer Sprache für die Analyse klassifiziert werden. Zum Test der Genderklassifikation wurden ca. 600 Accounts aus dem »#IchBinHanna«-Korpus manuell von zwei Mitarbeitenden des Pro-

jets klassifiziert. Die nach allen vier Schritten nicht klassifizierbaren Accounts waren vor allem Accounts ohne Klarnamen oder mit uneindeutigen Parametern, ohne Pronomen und mit uneindeutigen Namen. Da insgesamt zu wenig Accounts nicht binärer Personen für eine separate Auswertung im Sample vorhanden waren, wurden diese im Sinne des FLINTA\*-Begriffs mit den Frauen gemeinsam in der Kategorie Wissenschaftlerinnen\* zusammengefasst, um zumindest einen Hinweis auf intersektionale Besonderheiten und Ansatzpunkte für Folgeforschung geben zu können.<sup>5</sup>

## 2.2 Auswertung von Account-Bios und Tweet-Inhalten

Um uns der (Selbst-)Darstellung von Wissenschaftler\*innen auf Twitter zu nähern, wurden fünf Methoden aus dem Spektrum des Distant Reading angewendet. Zur Vorbereitung wurden die Daten einem Preprocessing unterzogen, in dessen Rahmen mittels fastText<sup>6</sup> die Sprache der Datensätze ermittelt wurde (Joulin et al. 2017). Sämtliche nicht deutsch- oder englischsprachigen Accounts wurden aus der Untersuchung ausgeschlossen. Alle Texte in beiden Sprachen wurden lemmatisiert (Montani et al. 2020; Wartena 2023). Pronomen-Angaben und Schreibweisen akademischer Grade (z. B. BA, PhD) wurden normalisiert. Füllwörter, URLs, Mastodon-Handles und Sonderzeichen wurden entfernt.

Mithilfe des Gender-Classifiers wurden nominale Personenreferenzen wie »Wissenschaftlerin« oder »Biologe« automatisch im Hinblick auf die enthaltene Genderinformation annotiert. Der Classifier enthält ein auf Wort- und Kontexteigenschaften basierendes probabilistisches CRF-Modell (siehe Manning et al. 2015), auf dessen Basis ausgewertet wird, ob eine Figurenreferenz im Text in eine der Kategorien »männlich«, »weiblich« oder »neutral« fällt. Der Classifier wurde für literarische Texte des 18. bis 21. Jahrhunderts optimiert und erreicht je nach Gattung und Genre eine Erkennungsgenauigkeit von 72–85 %. Da das Tool für eine von der hier betrachteten abweichende Domäne entwickelt wurde, kamen zusätzlich weitere Methoden zum Einsatz.

Um zu betrachten, ob Wissenschaftlerinnen\* in ihren Accounts und Tweets anders kommunizieren als Wissenschaftler, wurde mit Stilmetrie gearbeitet, die auf einem Abgleich von Wortfrequenzen beruht und u. a. zur Attribution von Autor\*innenschaft eingesetzt wird (Horstmann 2024 und 2018). Über die Oppose-Funktion des Tools Stylo<sup>7</sup> (Eder et al. 2016) wurden bevorzugte und vermiedene Wörter einer Autor\*innengruppe einer anderen gegenübergestellt. Um zu vermeiden, dass selbstevidente Ergebnisse die Analyse trüben (z. B. dass die Gruppe der Wissenschaftlerinnen\* »Historikerin« als Selbstbezeichnung gegenüber »Historiker« bevorzugt), wurde das Korpus für die Abfrage lemmatisiert.

Um herauszufinden, welche Wörter häufig in Verbindung mit den (Selbst-)Bezeichnungen stehen und somit zur (Selbst-)Beschreibung von Wissenschaftler\*innen dienen,

5 Die Autorinnen sind sich bewusst, dass nicht binäre Identitäten nicht mit weiblichen gleichzusetzen sind, und bedauern ausdrücklich, dass trotz des großformatigen Korpus die vorliegende Datenlage an dieser Stelle keine differenzierte Betrachtung ermöglicht.

6 <https://fasttext.cc/> (abgerufen am 27.06.2025).

7 <https://maciejeder.org/projects/stylo/> (abgerufen am 30.06.2025).

wurde (in Anlehnung an Schumacher und Flüh 2023) ein word2vec-Modell (Mikolov et al. 2013) trainiert und in Form einer Netzwerkanalyse visualisiert. Das word2vec-Verfahren ist eine Word-Embeddings-Methode, bei der Wörter in einem Korpus anhand ihrer Kontexte in Vektoren umgerechnet werden (siehe Schumacher 2024). Wörter, die häufig in einem ähnlichen Kontext stehen, haben ähnliche Vektoren (Schumacher 2024 und 2023), was – der distributionellen Hypothese der Linguistik folgend – auf Bedeutungsähnlichkeiten schließen lässt (Firth 1957). Das Vorgehen eignet sich besonders, um semantische Felder zu analysieren und zu verstehen, wie in einem bestimmten Korpus bestimmte Begriffe konzeptualisiert sind. Hierfür wurden die häufigsten 200 Wörter analysiert, sofern sie öfter als 10-mal (DE) bzw. 20-mal (EN) vorkamen (Stoppwörter<sup>8</sup> ausgeschlossen). Für das Tunen der Parameter wurden die Ergebnisse zunächst mit T-SNE (van der Maaten und Hinton 2008) visualisiert und dann mit der Software Gephi<sup>9</sup> als Netzwerkdiagramm dargestellt (Bastian et al. 2009). Das Korpus wurde jeweils in zwei Subkorpora unterteilt, um Wortfelder von Wissenschaftlerinnen\* und Wissenschaftlern vergleichend zu betrachten.

Ausgehend von den Ergebnissen der vergleichenden Studie von Egger und Yu (2022), die unterschiedliche Topic-Modeling-Verfahren für die Analyse von Twitter-Daten getestet haben, wurde für die Analyse der Tweets BERTopic<sup>10</sup> eingesetzt. BERTopic verwendet Word Embeddings, um Listen mit Wörtern zu erstellen, die innerhalb eines Korpus häufig zusammen vorkommen. Aus diesen semantischen Feldern können Rückschlüsse auf thematische Schwerpunkte von Textdaten gezogen werden. Für die Topic-Modeling-Analyse wurde der Gesamtdatensatz so reduziert, dass nur Tweets von Accounts erhalten blieben, die von Wissenschaftler\*innen, wissenschaftsnahen Personen, Forschungsinstitutionen und Universitäten stammen. Dieser Teildatensatz wurde außerdem so unterteilt, dass Vergleiche zwischen den Inhalten von Wissenschaftlerinnen\* und Wissenschaftlern möglich wurden.

### 3 Ergebnisse: Selbstdarstellung – Statistische Auswertung der Accounts

Eine statistische Auswertung der Accounts zeigt einige im Hinblick auf die (Selbst-)Darstellung von Forschenden relevante Trends. Zum einen werden besonders große Accounts ( $> 1.640$  Follower\*innen) oft von Professor\*innen geführt. Accounts mit Biografien auf Deutsch ( $m = 1.304,2$ ;  $sd = 2.761,7$ ) haben signifikant mehr Follower\*innen als auf Englisch geführte Accounts ( $t = 4,6663$ ;  $df = 5.255,5$ ;  $p < 0,001$ ;  $m = 1.083,6$ ;  $sd = 2.303,6$ ). In der englischsprachigen Gruppe posten Frauen ( $m = 7.482,4$ ;  $sd = 17.909$ ) signifikant weniger Tweets als Männer ( $t = 4,4752$ ;  $df = 17.452$ ;  $p < 0,01$ ; Männer:  $m = 8.730,5$ ;  $sd = 18.941$ ).

8 Als Stoppwörter gelten Wörter, die häufig vorkommen und nur wenig relevanten Informationswert besitzen, z.B. »und«, »die« oder »also«.

9 <https://gephi.org/> (abgerufen am 27.06.2025).

10 <https://github.com/MaartenGr/BERTopic> (abgerufen am 27.06.2025).

In beiden Sprachgruppen besteht ein Unterschied nach Gender. Bei den deutschsprachigen Accounts haben Wissenschaftlerinnen\* im Schnitt ( $m = 1.114,9$ ;  $sd = 2.361,5$ ) signifikant weniger Follower\*innen als Wissenschaftler ( $m = 1.433,6$ ;  $sd = 2.408,6$ ;  $t = 3,73$ ;  $df = 3.866,2$ ;  $p < 0,001$ ). Bei den englischsprachigen Accounts ist der Unterschied kleiner, aber immer noch signifikant ( $t = 2,7529$ ;  $df = 17,399$ ;  $p = 0,006$ ; Männer:  $m = 1.130,6$ ;  $sd = 2.998,2$ ; Frauen:  $m = 1.034,8$ ;  $sd = 2.188,4$ ).

## Aufbau von Account-Biografien

Biografien bestehen aus mehreren Bausteinen und werden in den untersuchten Accounts individuell gestaltet. Die sechs ausgewählten Bios (Abb. 2) sind typische, pseudonymisierte Beispiele. Den größten Raum nehmen fachspezifische Wörter ein, die die Expertise der Person abstecken (M1, M2, M3, W1, W2). Diese kann auch durch direkte Selbstbeschreibung als Expert\*in (M1, W2) oder indirekt durch die Selbstbezeichnung als (Fach-)Wissenschaftler\*in (M1, W1, W2), über akademische Positionen (M1, M3, W1, W2) oder Institutionen (M1, M3, W1, W2) markiert werden. Neben der professionalen Selbstdarstellung sind zudem oft Hobbys (Sportart (M2), »cat lover« (W2)) und Identitätsmarker (»she/her« (W1)) enthalten. Ebenso finden familiäre Rollen Erwähnung (M2, W3) oder Engagement, das mit der Fachdisziplin in Verbindung steht (W3).

Abb. 2: Typische Bausteine von Account-Biografien

 Professor für Denomination [Institution] Experte für Fachgebiet1a, Fachgebiet1b, Fachgebiet1c Mastodon: @name@mastodon.green (M1)	Trainer und Experte für Fachgebiet1a und Fachgebiet1b. Expertise1a, Expertise1b. Papa. Spielt Sportart1a und Sportart2b. (M2)	Wissenschaftler und freier Journalist. Fachgebiet1a   Fachgebiet1b   Fachgebiet1c Wissenschaftlicher Mitarbeiter @institut   Privater Account (M3)
(sie/ihr)  Juniorprofessorin für Denomination1a / Denomination1b [Institution]; Fachgebiet1a, Fachgebiet1b, Fachgebiet1c, Fachgebiet1d. (W1)	#hashtag @Institution Postdoctoral research fellow   Co-badged PhD in Fachgebiet1a   Fachgebiet1b expert   Cat lover (W2)	Wissenschaftlerin1a, Wissenschaftlerin1b, Mutter, Hobby, Optimistin, Engagement (W3)

## Rollenbezeichnungen

Im nächsten Schritt wurden die verwendeten Rollenbezeichnungen in der Selbstdarstellung genauer ausgewertet. Die mittels Classifier erzeugten 5.528 Annotationen der deutschsprachigen Account-Biografien umfassen 52,9 % männliche, 45,7 % weibliche

und 1,4 % neutrale Rollen. Die häufigsten Referenzen auf Frauen und Männer im Kontext von Twitter-Biografien sind in Abb. 3 dargestellt.

*Abb. 3: Häufigste Referenzen im Kontext der Biografien*

1. Platz Männer Historiker	4 Vater	5 Ingenieur	6 Student	1. Platz Frauen Mutter	4 Lehrerin	5 Mama
118 Vorkommen				107 Vorkommen		
2 Autor	82	76	71	2 Historikerin	64	57
7 Professor		9 Journalist	10 Experte	52		
105 3 Lehrer	71			79	7 Professore	37
8 Physiker				3 Autorin	41	10 Psycho- therapeutin
86	61	58	55	65	8 Soziologin	36

Einige Rollen tauchen in beiden Gruppen in derselben Reihenfolge auf, z. B. »Historiker\*in«, »Autor\*in«, »Lehrer\*in«. Bei Wissenschaftlern sind diese Rollen auf Platz 1 bis 3, bei Wissenschaftlerinnen\* steht allerdings »Mutter« an erster Stelle. Bei den Männern belegt »Vater« Platz 4. Bei Wissenschaftlern werden außerdem »Ingenieur«, »Physiker«, »Student«, »Journalist« und »Experte« oft verwendet, bei Wissenschaftlerinnen\* dagegen »Doktorandin«, »Soziologin«, »Mama« »Feministin« und »Psychotherapeutin«.

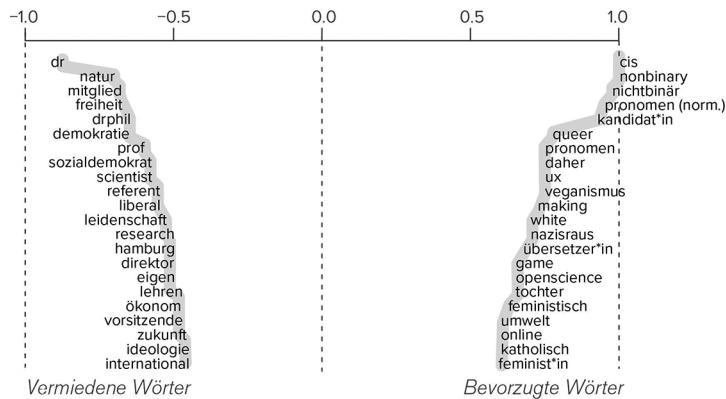
Beim Vergleich der Selbstbezeichnungen von Wissenschaftlern und Wissenschaftlerinnen\* fallen drei Dinge besonders auf: Zunächst einmal wird die Mutterrolle von Forscherinnen\* viel häufiger zur Selbstdarstellung genutzt als die Vaterrolle von Forschern. Nicht nur steht »Mutter« auf Platz 1 der meistgenannten Rollen in den Bios der Wissenschaftlerinnen\*, es kommt auf Platz 5 auch noch zusätzlich das persönlicher gefärbte Wort »Mama« vor. Eine zweite Auffälligkeit besteht darin, dass im Sample der von Männern geführten Accounts mehr Studenten vorkommen, während es bei den Frauen mehr Doktorandinnen gibt. Dies könnte darauf hindeuten, dass Männer früher in ihrer Karriere anfangen, sich auf Social Media im Forschungskontext darzustellen als Frauen. Bei beiden Samples fällt drittens auf, dass nicht nur direkt auf eine Beschäftigung in der Forschung hindeutende Berufsbezeichnungen vorkommen, sondern es bestimmte »forschungsnahe« Berufsgruppen zu geben scheint, die ebenfalls häufig mit den Hashtags »#Forschung« und »#PhDLife« twittern, wie etwa Autor\*innen, Lehrer\*innen, Journalist\*innen und Psychotherapeut\*innen.

## Verwendete Wörter

Vergleicht man die deutschsprachigen Account-Biografien, die Pronomen enthalten, mit solchen ohne Angabe, fällt ein starkes Ungleichgewicht in der Verteilung auf. In diesem Sample sind Wissenschaftler nur zu 29 % (n = 268) vertreten, während die übrigen 71 % (n = 611) FLINTA\* sind. Die Analyse der bevorzugten und vermiedenen Wörter zeigt nicht nur, dass diese Accounts als für die Genderthematik besonders sensibel eingestuft

werden können, sondern auch, dass sie zusätzlich stärker politische Inhalte posten. Bevorzugte Wörter wie »queer« oder »cis« belegen die Affinität zu Genderthemen.

*Abb. 4: Oppose-Abfrage für Biografien von Accounts, die Pronomen angeben (in Abgrenzung zu solchen, die das nicht tun)*



Werden in diesem verringerten Datensatz nun Accounts von Wissenschaftlerinnen\* solchen von Wissenschaftlern gegenübergestellt, zeigt sich, dass Wissenschaftlerinnen\* feministische und genderbezogene Wörter in der Selbstbeschreibung verwenden. Interessant ist aber auch, dass Wissenschaftskommunikation in den Posts dieser Gruppe eine besonders große Rolle spielt, wie sich in der Verwendung von Hashtags wie »#teamwissenschaft« und »#wisskomm« zeigt. Mit Wörtern wie »kreativ«, »pädagog\*in«, »kunst« zeigt sich außerdem ein geistes- und sozialwissenschaftlicher Schwerpunkt. Vermieden werden dagegen Wörter aus den Ingenieurs- und Technikwissenschaften sowie aus der Ökonomie (siehe Abb. 5). Hier zeigt sich bei den gendersensiblen Accounts eine geradezu überraschend starke Trennung von natur- und technikwissenschaftlichen Inhalten bei Männern und geisteswissenschaftlichen Schwerpunkten bei FLINTA\*.

Im englischsprachigen Datensatz zeigt sich ein ähnliches Bild: Bei Wissenschaftlern, die in der Account-Bio die englische Sprache nutzen, konzentrieren sich besonders viele Begriffe auf den Bereich von KI und Data Science, außerdem benutzen sie öfter Begriffe aus der Wirtschaft (»economist«, »entrepreneur«) und solche, die eine Karrierestufe markieren (»ceo«, »vorsitz«). Bei Wissenschaftlerinnen\* finden sich erneut Begriffe wie »feminism«, »queer«, »intersectional« und »gender« sowie mit »museum« oder »anthropologist« auch solche, die auf soziale und kulturelle Schwerpunkte hindeuten. Auch Wissenschaftskommunikation (»sicom«) kommt vor. Im Gegensatz zum deutschsprachigen Datensatz sind aber auch Hinweise auf Naturwissenschaften und Technik (»stem«) enthalten. Bei der Deutung dieser Daten muss berücksichtigt werden, dass es sich lediglich um eine sprachliche Differenzierung handelt, da viele Forschende aus dem deutschsprachigen Raum ihre Biografien auf Englisch verfassen. Eventuell zeigt sich hier eine Forschungskultur der MINT-Fächer, die dazu führt, dass Wissenschaftlerinnen\* aus diesem Bereich besonders häufig englischsprachige Bios haben.

Abb. 5: Oppose-Abfrage für Biografien von deutschsprachigen Accounts von Wissenschaftlerinnen\* (in Abgrenzung zu Wissenschaftlern)

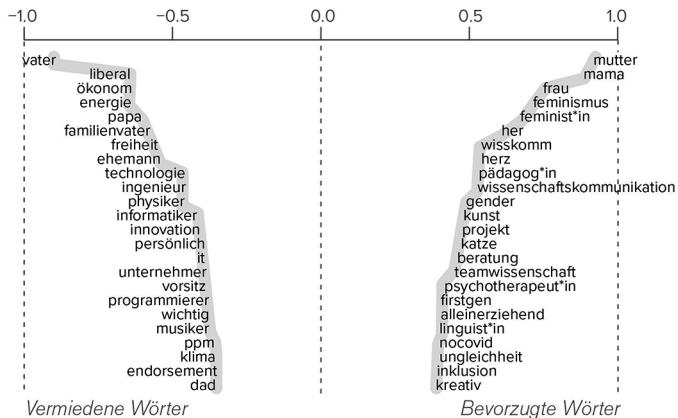
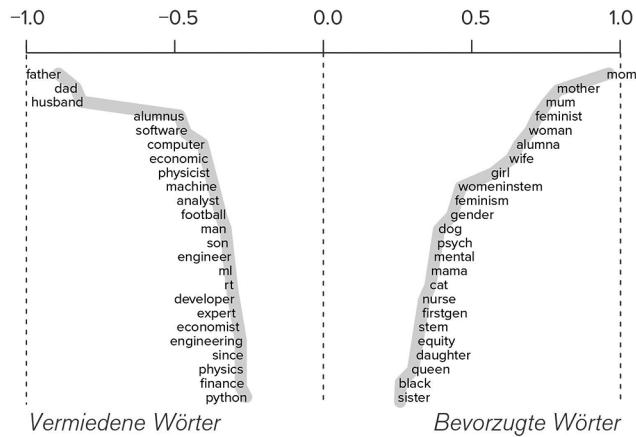


Abb. 6: Oppose-Abfrage für Biografien von englischsprachigen Accounts von Wissenschaftlerinnen\* (in Abgrenzung zu Wissenschaftlern)

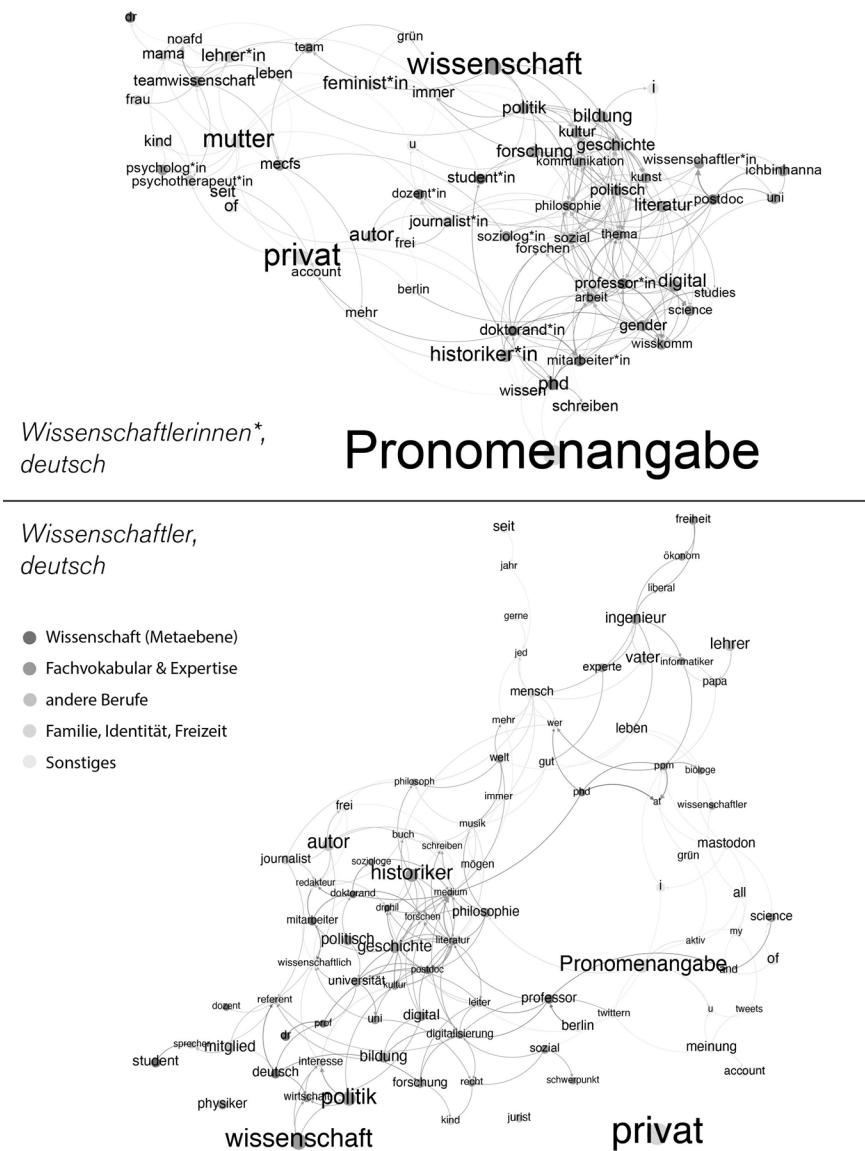


## Word Embeddings

Aufbauend auf den Ergebnissen der Word Embeddings wurden Netzwerkdiagramme für beide Gruppen in beiden Sprachen erstellt (Abb. 7).

Je breiter die Verbindung zwischen zwei Wörtern, desto ähnlicher sind deren häufigste Kontexte. Die Darstellung clustert Wörter in verschiedene Wortgruppen, beispielsweise Berufe, Verben, Fachvokabular und Familie.

Abb. 7: Semantische Felder in Account-Biografien von Wissenschaftlerinnen\* und Wissenschaftlern



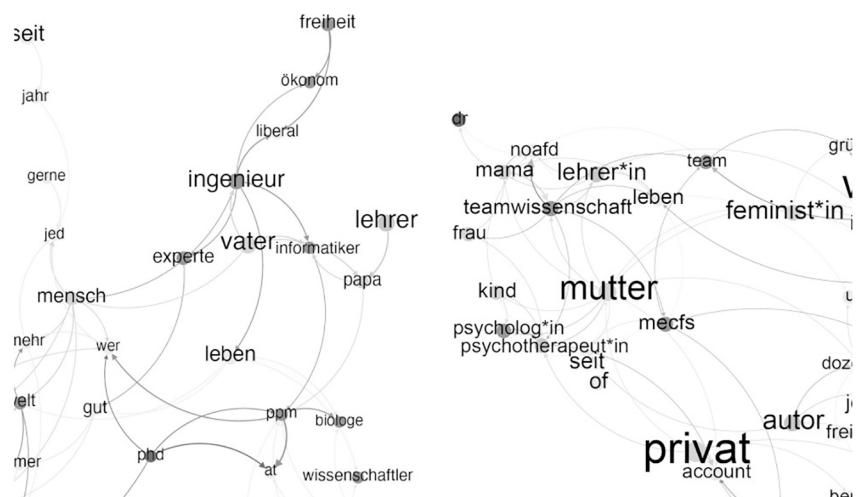
Es zeigt sich, dass der Großteil der Wortfelder Wörter mit Bezug zu den verschiedenen Fachdisziplinen sind, die sich zwischen Gendergruppen und Sprachen unterschiedlich gewichten. In allen vier Gruppen bilden sich große Cluster mit Fachvokabular, Expertisefeldern und wissenschaftlichen Berufsbezeichnungen. Hier werden sprachliche Unterschiede sichtbar. Im Englischen gibt es eine größere Bandbreite an fachlichen Begriffen, da diese im Gegensatz zum Deutschen nicht zusammengesetzt werden: Wäh-

rend z. B. »Kunstgeschichte« im deutschsprachigen Korpus zu selten ist, als dass sie im Netzwerk vorkommt, sind im Englischen sowohl »art« als auch »history« als alleinstehende Begriffe häufig genug, um ausgewertet werden zu können.

Neben den vielen fachspezifischen Begriffen finden sich im Datensatz auch Wortsfelder, die Privatheit ausdrücken. Im englischsprachigen Korpus sind das etwa »love«, »music« oder »nature«. Vergleicht man im Deutschen die Elternrollen, zeigt sich, dass »mutter« und »mama« mit deutlich mehr anderen Rollenbeschreibungen verknüpft sind, also ähnlich verwendet werden, »vater« und »papa« dagegen nur wenige Verknüpfungen zu Rollenbeschreibungen haben. Abb. 8 zeigt Ausschnitte, in denen Wortverknüpfungen der Elternrollen aufgezeigt werden. »Vater« ist beispielsweise mit den Wörtern »mensch«, »ingenieur«, »informatiker« und »papa« verknüpft, während »mutter« mit »mama«, »teamwissenschaft«, »lehrer\*in«, »feminist\*in«, »dozent\*in«, »immer«, »seit«, »kind«, »psycholog\*in«, »account« und »psychotherapeut\*in« verbunden ist. Es zeigt sich somit eine engere Einflechtung der Mutterschaft in andere Kontexte, die im Englischen nicht gegeben ist.

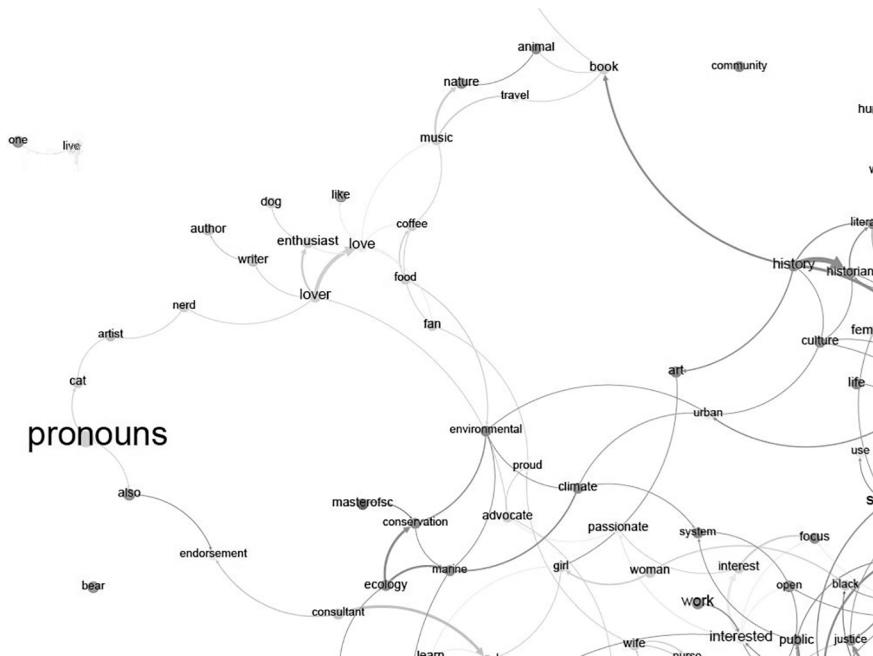
Wissenschaftler setzen häufiger Rahmenbedingungen für ihren Account, indem sie die Meinungen, die sie auf dem Account teilen, als privat eingrenzen. »Privat« ist das am häufigsten verwendete Wort der deutschsprachigen Wissenschaftler. Beispiele für Verwendungskontexte sind »Hier privat unterwegs« und »Äußerungen sind meine private Meinung«. Im englischsprachigen Korpus ist hierfür die Formulierung »views are my own« geläufig.

Abb. 8: Häufige Kontexte von Elternrollen (jeweils Ausschnitt; links: M, deutsch; rechts: W, deutsch)



Während bei deutschsprachigen Wissenschaftlerinnen\* bis auf die Mutterschaft kaum andere private Identitätsmerkmale oder Freizeitaktivitäten vorkommen, bildet sich bei den englischsprachigen Wissenschaftlerinnen\* ein Cluster aus Hobbys und privaten Interessen, die sich mit Identitätsmarkern und disziplinspezifischen Wörtern verknüpfen. Auch »wife« kommt vor – unabhängig vom Elternkontext.

Abb. 9: Cluster privater Identitätsmerkmale in englischsprachigen Account-Biografien (W, Ausschnitt)



Auffällig ist außerdem, dass nicht akademische Bezeichnungen der Karrierestufe in der deutschen Sprache bei Frauen nicht vorkommen. Bei deutschsprachigen Wissenschaftlern wird das Wort »leiter« oft verwendet, im Zusammenhang mit »professor«, »drphil«, »postdoc« und fachspezifischen Wörtern. Im englischsprachigen Bereich tauchen Wörter wie »head«, »chair« oder »leader« bei allen Gendern auf. Ein Unterschied ist die Einbettung dieser Wörter. Bei Männern bildet sich ein Cluster in Verknüpfung mit der Professur, Business und der Universität (Abb. 10). Bei Frauen dagegen gibt es keine Verknüpfung zur Professur, sondern nur mit der Universität und ökonomischen Begriffen (Abb. 11). Das gilt insbesondere für die Rolle »director«.

Während sich 72 Wissenschaftler als »experte« bezeichnen, verwenden Wissenschaftlerinnen\* das Wort deutlich seltener zur Eigenbeschreibung ( $n = 21$ ). Im englischsprachigen Korpus sind es 116 Wissenschaftler und 59 Wissenschaftlerinnen\*, die diese Rollenbeschreibung verwenden. In beiden Sprachen ist das Wort bei den Wissenschaftlerinnen\* somit nicht unter den 200 am häufigsten vorkommenden Wörtern. Im deutschsprachigen Teil wird es oft im Zusammenhang mit »ingenieur« verwendet (siehe

Abb. 8), im englischsprachigen ist es eher im Kontext von »specialist« und »analyst« zu finden (Abb. 12).

Abb. 10: Cluster beruflicher Rollen in englischsprachigen Account-Biografien (M, Ausschnitt)

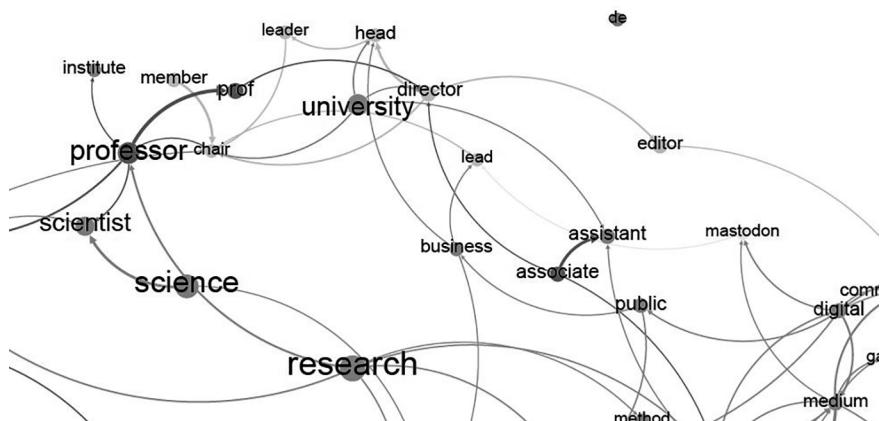


Abb. 11: Cluster beruflicher Rollen in englischsprachigen Account-Biografien (W, Ausschnitt)

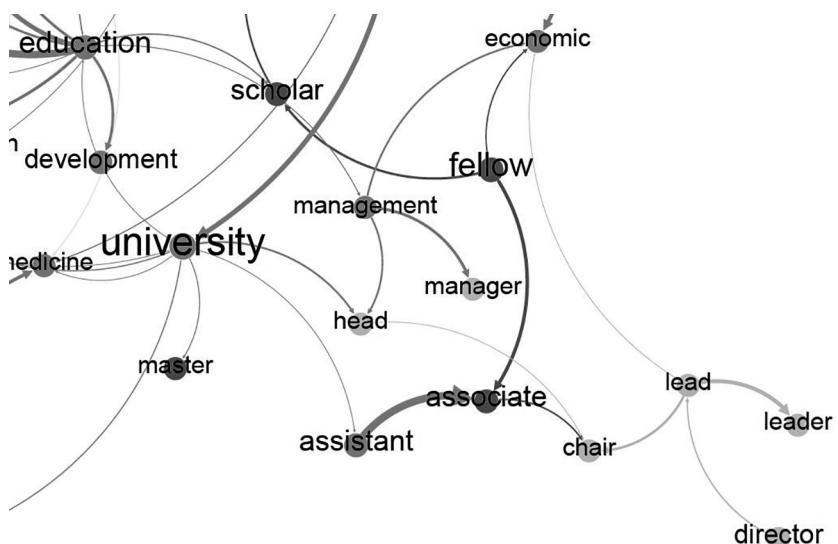
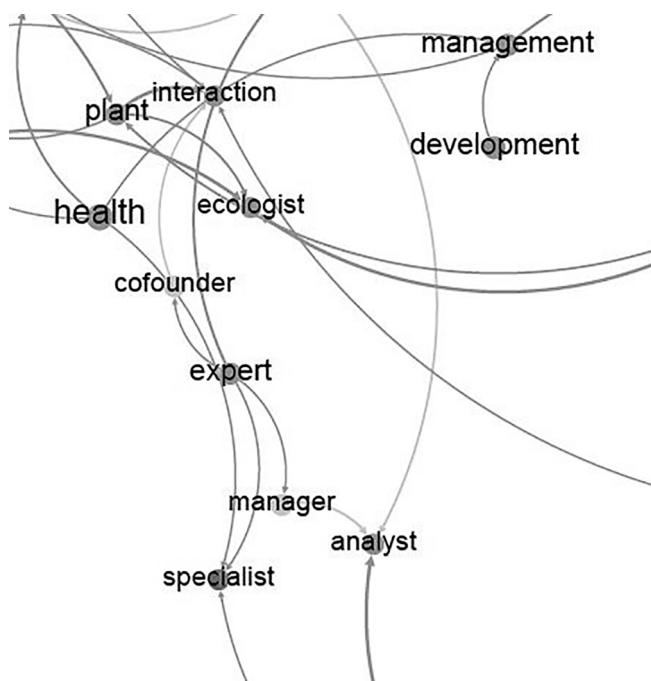


Abb. 12: Expert-Cluster in englischsprachigen Account-Biografien (M, Ausschnitt)



## Topic Modeling

Für das Topic Modeling der Tweets von Wissenschaftler\*innen, wissenschaftsnahen Personen und wissenschaftlichen Accounts wurde zuerst ein BERTopic-Modell mit 100 Topics erstellt und im Anschluss der Umfang auf 30 Topics reduziert. Dabei konnten insgesamt 1.121 von 3.749 Tweets als Ausreißer eingeordnet werden. Für eine bessere Darstellung wurden für die acht häufigsten Topics auf Basis der repräsentativen Wörter und Tweets Oberbegriffe formuliert (Tab. 1).

Die Resultate zeigen, dass sich Tweets von Wissenschafts-Accounts vorrangig mit wissenschaftsinternen Themen wie Forschung, Arbeitsbedingungen und digitaler Lehre auseinandersetzen. Zudem konnten politische und gesellschaftliche Themen wie ME/CFS, Long Covid, die Energiekrise oder Klimaaktivismus identifiziert werden. Bei der Differenzierung nach Tweet-Kategorien hat sich gezeigt, dass die Retweets am häufigsten das Thema 4 (Wissenschaftsstandort Berlin und Bewerbungen) enthalten, was möglicherweise auf Tweets zu Stellenausschreibungen, die das Wort »Bewerbung« beinhalten, zurückzuführen sein könnte. In den originären Tweets ist Thema 1 (Forschung und Arbeitsbedingungen in der Wissenschaft) am präsentesten.

Um gegenüberzustellen, welche thematischen Schwerpunkte Wissenschaftler\*innen in ihren Tweets nutzen und worüber Wissenschaftler häufiger kommunizieren, wurde untersucht, inwiefern sich die Verteilung der häufigsten sechs Topics in den

Tweets nach Gender unterscheidet (Abb. 13). Während der Abfrage wurden neue Topics modelliert, weshalb die jeweiligen repräsentativen Wortlisten mit angegeben sind.

Die Wortlisten zeigen, dass die verschiedenen Topics nicht nur unterschiedliche Prozentsätze einnehmen, sondern auch unterschiedliche Schwerpunkte setzen. Forscherinnen\* wie Forscher schreiben in rund 40 % ihrer Tweets mit den Hashtags »#Forschung« und »#PhDLife« über Forschung und Arbeitsbedingungen in der Wissenschaft. Während bei den Wissenschaftlerinnen\* auch Lehre in die Top-Topics gehört, ist das bei Wissenschaftlern nicht der Fall. Im zweiten Topic kommt bei beiden Gruppen »Lehre« vor und dies häufig im Kontext von Digitalität (»digitale«, »ChatGPT«). Bei Forschern ist allerdings auch die Titelbezeichnung »prof« Teil dieses Topics, bei Forscherinnen\* hingegen der positive Ausdruck »freuen«.

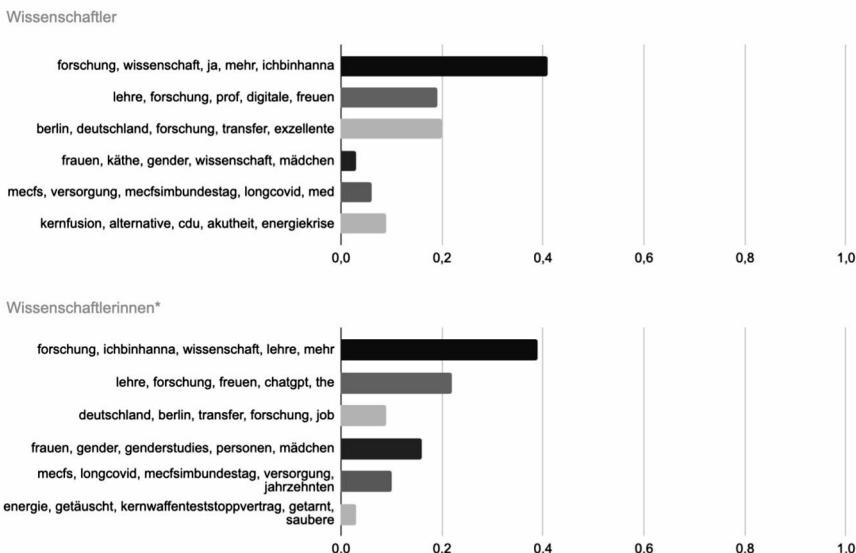
*Tab. 1: Die acht häufigsten Tweet-Topics der Wissenschaftler\*innen*

Nr.	Thema	Repräsentative Wörter	Anzahl Tweets
1	Forschung und Arbeitsbedingungen in Wissenschaft	[>forschung<, >ichbinhanna<, >wissenschaft<, >lehre<, >immer<, >ja<, >mehr<, >deutschland<, >gibt<, >ermöglicht<]	736
2	Digitale Lehre und Forschung	[>forschung<, >digitale<, >the<, >lehre<, >thema<, >chatgpt<, >digitalen<, >podcast<, >fdm<, >digitalisierung<]	328
3	ME/CFS und Long Covid	[>mecfs<, >versorgung<, >mecfsimbundestag<, >forschung<, >longcovid<, >mehr<, >gibt<, >menschen<, >jahrzehnten<, >helfen<]	294
4	Wissenschaftsstandort Berlin und Bewerbungen	[>berlin<, >freuen<, >bewerbungen<, >exzellente<, >heimatort<, >brain<, >wusstet<, >istberlin<, >wirberlin<, >klimaneutralen<]	164
5	Covid und Endemie	[>heißt<, >longcovid<, >covid<, >viren<, >prävention<, >endemisch<, >respiratorischen<, >endemie<, >erreger<, >therapeutika<]	141
6	Energiekrise	[>kernfusion<, >wasserstoff<, >alternative<, >energie<, >fusion<, >alten<, >akutheit<, >fusionshype<, >grünen<, >akw<]	120
7	Klimaaktivismus	[>geht<, >klima<, >engagiert<, >ehren<, >fleissig<, >klimaschützer<, >ansinnen<, >demonstrieren<, >demonstriert<, >demos<]	113
8	Gender Studies und Wissenschaftlerinnen	[>frauen<, >genderstudies<, >mädchen<, >wissenschaftlerinnen<, >wissenschaft<, >weise<, >womeninscience<, >personen<, >>tag<, >internationalen<]	106

Die größten genderspezifischen Unterschiede zeigen sich bei den Topics »Gender Studies und Wissenschaftlerinnen« sowie »Energiekrise«. Während Frauen in der Wissenschaft in 16 % ihrer Tweets (91 von 558 originären Tweets) Genderthemen zum Inhalt machen, kommen diese nur in 3 % (28 von 849 originären Tweets) der Tweets von Män-

nern vor. Andersherum geht es in rund 9 % der Tweets von Männern (91 von 849 originären Tweets) in der Wissenschaft um die Energiekrise und verwandte Themen, während Frauen nur in 3 % ihrer Tweets (19 von 558 originären Tweets) darüber kommunizieren. Hier zeigt sich erneut die Tendenz der oben stehenden Auswertungen: Wissenschaftlerinnen\* stellen in ihren Tweets das geistes- und kulturwissenschaftliche Thema »Gender« häufiger in den Mittelpunkt, während Wissenschaftler öfter über das naturwissenschaftliche Thema »Klima« twittieren. Die Analyse zeigt allerdings auch, dass trotz unterschiedlicher Ausprägung alle Forschenden während des Untersuchungszeitraums viel über Forschung und Bedingungen in der Wissenschaft, über Lehre und die Corona-krise kommuniziert haben.

*Abb. 13: Sechs häufigste Topics und ihre repräsentativsten Wörter nach Gender*



## 4 Diskussion der Ergebnisse

Im Datensatz sind Wissenschaftler\*innen diverser Fachbereiche vertreten, mit besonderer Häufung der Geistes-, Sozial- und Politikwissenschaften. Das deckt sich mit Beobachtungen von Ke et al. (2017), die ermittelt haben, dass diese Gruppe besonders oft soziale und politische Inhalte teilt und daher auch verstärkt auf Twitter vertreten ist. Bei den FLINTA\* fallen zudem viele Wörter auf, die auf sozial- und kulturwissenschaftliche Interessen hindeuten und/oder in den Bereich der Antidiskriminierung fallen. In der Gruppe der Wissenschaftler sind dagegen Wörter aus den MINT-Fächern und den Wirtschaftsdisziplinen präsent. Diese Unterschiede könnten mit den verschiedenen Geschlechterverhältnissen in den Disziplinen der jeweiligen Wissenschaftler\*innen zusammenhängen. Die Ergebnisse deuten aber auch auf eine gesamtgesellschaftliche

Prägung der Geistes- und Sozialwissenschaften als stärker weiblich und der Natur- und Technikwissenschaften als stärker männlich konnotiert hin. Interessant ist hier auch, dass die Trennschärfe im englischsprachigen Datensatz geringer ist. Insgesamt zeigt sich, dass sich Wissenschaftler\*innen über Gendergrenzen hinweg größtenteils professionell positionieren. Die Selbstbezeichnungen, die am häufigsten in Account-Bios genutzt werden, stammen aus wissenschaftsnahen Kontexten. Viele der Accounts in unserem Datensatz stammen außerdem von Early-Career-Wissenschaftler\*innen, was sich durch die Präsenz von Wörtern wie »phD«, »postdoc« und »student« zeigt. Bei deutschsprachigen Wissenschaftlerinnen\* kommen Wörter wie »leiterin« nicht unter den 200 häufigsten Begriffen vor, bei deutschsprachigen Wissenschaftlern hingegen schon, vor allem im Zusammenhang mit höheren akademischen Positionen. Ein Grund dafür könnte das Phänomen der *Leaky Pipeline* sein, das ausdrückt, dass mit steigender Karrierestufe der Frauenanteil sinkt (Konsortium Bundesbericht wissenschaftlicher Nachwuchs 2021: 106). Ebenso denkbar wäre, dass Wissenschaftlerinnen\* seltener ihren Expertinnenstatus zur Selbstdarstellung nutzen, denn auch das Wort »experte« wird deutlich häufiger von den männlichen Kollegen gebraucht. Dies könnte zum einen daran liegen, dass Männer in klassischen Medien überdurchschnittlich oft extern als Experten markiert werden (MaLisa Stiftung 2020), aber eventuell auch an ihrer Selbsteinschätzung, die durch die Wissenschaftsstruktur entstehen könnte. FLINTA\* nutzen zudem häufiger Hashtags wie »#wisskomm«, was ein Hinweis darauf sein könnte, dass sie ihre Tweets besonders legitimieren möchten. Zukünftige Forschung könnte die Gründe für die (fehlende) Selbstbezeichnung als Expert\*in genauer untersuchen.

In unserem Datensatz zeigt sich, dass beide Gruppen sich nicht nur im professionellen Kontext beschreiben, sondern in der Biografie auch private Aspekte vorkommen. Während Wissenschaftler ihr Profil oft explizit mit »privater Account« kennzeichnen, nennen weiblich gelesene und diverse Menschen oft ihre Elternrolle. Besonders die Rolle der Mutter scheint ein wichtiger Teil der Selbstdarstellung zu sein. Möglicherweise steht hier die Idee im Vordergrund, Mutterschaft als unbezahlte Care-Arbeit und wichtigen Teil der Identitätskonstruktion abzubilden. Bei englischsprachigen Accounts besteht dieser Zusammenhang nicht. Dass die Selbstdarstellung als Elternteil nicht in direktem Zusammenhang mit einer höheren Follower\*innenzahl steht, deckt sich mit den Erkenntnissen von Zhang und Lu (2023), die zeigen, dass private Selbstdarstellung bei Social-Media-Nutzung von Wissenschaftler\*innen nicht zu einer höher wahrgenommenen Authentizität und/oder stärkeren Bindung zu Follower\*innen führt, sondern dafür eher die professionelle Selbstdarstellung verantwortlich ist.

Dass das Wort »privat« bei Wissenschaftlern mit deutschsprachigen Accounts so präsent ist, steht allerdings nicht in Verbindung mit einer Privatheit hinsichtlich einer Elternrolle, sondern in Zusammenhang mit der Einordnung der eigenen Tweets als nicht institutionelle Meinung. Es dient somit der Abgrenzung der professionellen von der privaten Persona (Hanusch und Bruns 2017). Die Daten von Wissenschaftlerinnen\* zeigen hingegen eher eine Kontextualisierung im (nicht privaten) Forschungsdiskurs. Tweets werden über Hashtags wie »#wisskomm« in einen professionellen Zusammenhang gestellt. Unsere Ergebnisse stehen hier im Gegensatz zur erforschten Selbstdarstellung von Journalist\*innen auf Social Media, bei denen öfter Frauen diese Art von Disclaimer verwenden (Hanusch und Bruns 2017).

Im Rahmen der textorientierten und auf Gender fokussierten Distant-Reading-Analyse konnte festgestellt werden, dass insgesamt nur ein relativ geringer Anteil der Twitter-Accounts mit Pronomenangaben versehen ist. Die Auswertung ergab, dass diese Accounts auch inhaltlich das Thema Gender aufgreifen und als besonders sensibel im Umgang mit Genderthemen beschrieben werden können. Ein sehr kleiner Anteil der Accounts wurde über Nennung entsprechender Pronomen als von nonbinären Personen geführt ausgewiesen. Auch wenn die Anzahl dieser Accounts zu klein war, um als eigene Gruppe in dieser vergleichenden Analyse berücksichtigt zu werden, konnten durch die Berücksichtigung und Annotation dieser Accounts erste Einblicke in intersektionale Dimensionen der (Selbst-)Darstellung von Forschenden auf Twitter gegeben und ein Ausgangspunkt für weitere qualitativ-inhaltliche Analysen geschaffen werden.

## Literatur

- Amarasekara, Inoka und Will J. Grant. 2019. Exploring the YouTube Science Communication Gender Gap: A Sentiment Analysis. *Public Understanding of Science* 28, Nr. 1 (Januar): 68–84. <https://doi.org/10.1177/0963662518786654>.
- Bastian, Mathieu, Sébastien Heymann und Jacomy Mathieu. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the Third International AAAI Conference on Weblogs and Social Media* 3, Nr. 1 (20. Mai): 361–362. <https://doi.org/10.1609/icwsm.v3i1.13937>.
- Bowman, Timothy D. 2015. Differences in Personal and Professional Tweets of Scholars. *arXiv:1505.01515* (6. Mai). <https://doi.org/10.48550/arXiv.1505.01515>.
- Butler, Judith. 2003. *Das Unbehagen der Geschlechter*. Gender Studies: Edition Suhrkamp, Bd. 722. Frankfurt a.M.: Suhrkamp.
- Cheng, Jonathan. 2020. Fleshting Out Models of Gender in English-Language Novels (1850–2000). *Journal of Cultural Analytics* 5, Nr. 1 (29. Januar). <https://doi.org/10.22148/001c.11652>.
- Cheplygina, Veronika, Felienne Hermans, Casper Albers, Natalia Bielczyk und Ionica Smeets. 2020. Ten simple rules for getting started on Twitter as a scientist. *PLoS Computational Biology* 16, Nr. 2 (10. Februar): e1007513. <https://doi.org/10.1371/journal.pcbi.1007513>.
- Collins, Kimberly, David Shiffmann und Jenny Rock. 2016. How Are Scientists Using Social Media in the Workplace? *PLoS ONE* 11, Nr. 10 (12. Oktober): e0162680. <https://doi.org/10.1371/journal.pone.0162680>.
- Connell, Raewyn W. 1996. *Gender and power society, the person and sexual politics*. Cambridge: Polity Press.
- Côté, Isabelle M. und Emily S. Darling. 2018. Scientists on Twitter: Preaching to the choir or singing from the rooftops? *FACETS* 3, Nr. 1 (28. Juni): 682–694. <https://doi.org/10.1139/facets-2018-0002>.
- de Beauvoir, Simone. 1992. *Das andere Geschlecht. Sitte und Sexus der Frau*. Neuübersetzung. Reinbek: Rowohlt.

- Eder, Maciej, Jan Rybicki und Mike Kestemont. 2016. Stylometry with R: A Package for Computational Text Analysis. *The R Journal* 8, Nr. 1 (22. Dezember): 107–121. <https://doi.org/10.32614/RJ-2016-007>.
- Egger, Roman und Joanne Yu. 2022. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology* 7 (6. Mai): p.886498. <https://doi.org/10.3389/fsoc.2022.886498>.
- Eysenbach, Gunther. 2011. Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact. *Journal of Medical Internet Research* 13, Nr. 4 (Oktober–Dezember): e123. <https://doi.org/10.2196/jmir.2012>.
- Firth, John R. 1957. A synopsis of linguistic theory 1930–55. In: *Studies in Linguistic Analysis*, hg. von John R. Firth. Oxford: Blackwell, 1–32.
- Freitas, Cláudia und Diana Santos. 2023. Human Depiction in Portuguese. Distant Reading Brazilian and Portuguese literature. *Journal of Computational Literary Studies* 2, Nr. 1 (11. Dezember): 1–29. <https://doi.org/10.48694/jcls.3576>.
- Hanusch, Folker und Axel Bruns. 2017. Journalistic Branding on Twitter: A representative study of Australian journalists' profile descriptions. *Digital Journalism* 5, Nr. 1: 26–43. <https://doi.org/10.1080/21670811.2016.1152161>.
- Hennig, Anne und Sarah Kohler. 2020. Einflussfaktoren bei der Social-Media-Nutzung in der Wissenschaftskommunikation. *Publizistik* 65, Nr. 4 (15. Oktober): 593–615. <https://doi.org/10.1007/s11616-020-00618-z>.
- Horstmann, Jan. 2024. Methodenbeitrag: Stilometrie. *forTEXT* 1, Nr. 1 (26. Februar): 1–7. <https://doi.org/10.48694/fortext.3769>.
- Joulin, Armand, Edouard Grave, Piotr Bojanowski und Tomas Mikolov. 2017. Bag of Tricks for Efficient Text Classification. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, hg. von Mirella Lapata, Phil Blunsom und Alexander Koller. April. 427–431. <https://aclanthology.org/E17-2068.pdf> (zugegriffen: 30 Juni 2025).
- Jünger, Jakob und Birte Fähnrich. 2020. Does really no one care? Analyzing the public engagement of communication scientists on Twitter. *New Media & Society* 22, Nr. 3 (März): 387–407. <https://doi.org/10.1177/1461444819863413>.
- Ke, Qing, Yong-Yeol Ahn und Cassidy R. Sugimoto. 2017. A systematic identification and analysis of scientists on Twitter. *PLoS ONE* 12, Nr. 4 (11. April): e0175368. <https://doi.org/10.1371/journal.pone.0175368>.
- Klar, Samara, Yanna Krupnikov, John B. Ryan, Kathleen Searles und Yotam Shmargad. 2020. Using social media to promote academic research: Identifying the benefits of twitter for sharing academic work. *PLoS ONE* 15, Nr. 4 (6. April): e0229446. <https://doi.org/10.1371/journal.pone.0229446>.
- Konsortium Bundesbericht Wissenschaftlicher Nachwuchs (Hg.). 2021. *Bundesbericht wissenschaftlicher Nachwuchs 2021. Statistische Daten und Forschungsbefunde zu Promovierenden und Promovierten in Deutschland*. Bielefeld: wbv Media GmbH & Co. KG. <https://doi.org/10.3278/6004603aw>.
- Lamb, Clayton T., Sophie L. Gilbert und Adam T. Ford. 2018. Tweet success? Scientific communication correlates with increased citations in Ecology and Conservation. *PeerJ* 6 (12. April): e4564. <https://doi.org/10.7717/peerj.4564>.

- Laser, Stefan, David Adler, Migle Bareikyte, Elen Le Foll, Christian Funk und Jana Lasser. 2021. #IchBinHanna Research Collective. Analysis of Tweets for the #IchBinHanna Campaign. <https://github.com/LaserStefan/IchbinHanna> (zugegriffen: 27. Juni 2025).
- MaLisa Stiftung. 2020. Wer wird gefragt? Geschlechterverteilung in der Corona-Berichterstattung. Zentrale Ergebnisse einer Analyse zur Geschlechterverteilung in der Corona-Berichterstattung im Fernsehen und in Online-Auftritten deutscher Printmedien im Auftrag der MaLisa Stiftung. Mai. [https://static1.squarespace.com/static/672b6a7044080f328d01a3f9/t/6762ab2995ff876cc8abce69/1734519593865/Studie\\_MaLisa\\_Geschlechterverteilung\\_in\\_der\\_Corona\\_Berichterstattung.pdf](https://static1.squarespace.com/static/672b6a7044080f328d01a3f9/t/6762ab2995ff876cc8abce69/1734519593865/Studie_MaLisa_Geschlechterverteilung_in_der_Corona_Berichterstattung.pdf) (zugegriffen: 2. Juli 2025).
- Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard und David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In: *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Baltimore (MD): Association for Computational Linguistics, 55–60. <https://aclanthology.org/P14-5010.pdf> (zugegriffen: 30. Juni 2025).
- Mikolov, Tomas, Kai Chen, Greg Corrado und Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *arXiv:1301.3781* (7. September). <https://doi.org/10.48550/arXiv.1301.3781>.
- Mohammadi, Ehsan, Mike Thelwall, Mary Kwasny und Kristi L. Holmes. 2018. Academic information on Twitter: A user survey. *PLoS ONE* 13, Nr. 5 (17. Mai): e0197265. <https://doi.org/10.1371/journal.pone.0197265>.
- Montani, Ines, Matthew Honnibal, Adriane Boyd, Sophie Van Landeghem und Henning Peters. 2020. spaCy: Industrial-strength Natural Language Processing in Python (Version 1.2.0) [Software]. [doi.org/10.5281/zenodo.1212303](https://doi.org/10.5281/zenodo.1212303).
- Phillips, Alexandra A., Catherine R. Walsh, Korie A. Grayson, Camilla E. Penney und Fatima Husain. 2022. Diversifying Representations of Female Scientists on Social Media: A Case Study From the Women Doing Science Instagram. *Social Media + Society* 8, Nr. 3 (29. Juli): 1–17. <https://doi.org/10.1177/20563051221113068>.
- Ricoeur, Paul. 1987. Narrative Identität. In: *Heidelberger Jahrbücher*, hg. von Elmar Mittler. Berlin/Heidelberg: Springer, 57–67. [https://doi.org/10.1007/978-3-642-71777-2\\_5](https://doi.org/10.1007/978-3-642-71777-2_5).
- Rybicki, Jan. 2016. Vive la différence: Tracing the (authorial) gender signal by multivariate analysis of word frequencies. *Digital Scholarship in the Humanities* 31, Nr. 4 (Dezember): 746–761. <https://doi.org/10.1093/llc/fqv023>.
- Schumacher, Mareike und Marie Flüh. 2023. Made to Be a Woman: A case study on the categorization of gender using an individuation-based approach in the analysis of literary texts. *Digital Humanities Quarterly* 17, Nr. 3: <https://www.digitalhumanities.org/dhq/vol/17/3/000728/000728.html> (zugegriffen: 30. Juni 2025).
- Schumacher, Mareike. 2021. *StanfordNER Gender-Classifier* (Version 1.0) [Software]. Zenodo. Oktober. <https://doi.org/10.5281/zenodo.5555952>.
- Schumacher, Mareike. 2024. Methodenbeitrag: word2vec. *forTEXT* 1, Nr. 10 (30 Oktober): 1–12. <https://doi.org/10.48694/fortext.3815>.
- Underwood, Ted, David Bamman und Sabrina Lee. 2018. The Transformation of Gender in English-Language Fiction. *Journal of Cultural Analytics* 3, Nr. 2 (13. Februar): 1–25. <https://doi.org/10.22148/16.019>.

- van der Maaten, Laurens und Geoffrey Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9, Nr. 86 (November): 2579–2606. <https://jmlr.org/pers/volume9/vandermaaten08a/vandermaaten08a.pdf> (zugegriffen: 30. Juni 2025).
- Vianne, Laurine, Yoann Dupont und Jean Barré. 2023. Gender Bias in French Literature. *CHR Computational Humanities Research* 2023, Vol-3558: 247–262. Proceedings of the Computational Humanities Research Conference 2023. Paris, 6.–8. Dezember. <https://ceur-ws.org/Vol-3558/paper2449.pdf> (zugegriffen: 30. Juni 2025).
- Wartena, Christian. 2023. The Hanover Tagger – Lemmatization, Morphological Analysis and POS Tagging in Python (Version 1.1.0) [Software]. <https://doi.org/10.25968/opus-2457>.
- Zhang, Annie L. und Hang Lu. 2023. Scientists as Influencers: The Role of Source Identity, Self-Disclosure, and Anti-Intellectualism in Science Communication on Social Media. *Social Media + Society* 9, Nr. 2 (22. Juni): 1–16. <https://doi.org/10.1177/20563051231180623>.