jfml Journal für Medienlinguistik
Journal for Media Linguistics

Vol 8 (2026), No 1: 91–123
DOI: 10.21248/jfml.2026.79

Gutachten und Kommentare unter:
https://dp.jfml.org/2025/opr-flausch/

Flausch

Unterstützende Sprachhandlungen in sozialen Medien

Yulia Clausen & Tatjana Scheffler

Abstract

In virtuellen Räumen finden vielfältige Formen sprachlicher und nicht-sprachlicher Interaktion statt. Im Vergleich zu Hassrede wurde positive, unterstützende Sprache bislang weniger untersucht. Diese Studie widmet sich daher dem Phänomen „Flausch“, das als kommunikative Handlung verstanden wird, die das soziale Selbstbild der Adressat*innen („Face“; Goffman 1967) stärkt und soziale Nähe herstellt. Zur systematischen Analyse entwickeln wir eine Typologie der Flauschausdrücke mit sieben Hauptkategorien: *positive feedback*, *affection declaration*, *compliment*, *encouragement*, *gratitude*, *group membership* und *sympathy*. Diese werden in über 46 000 deutschsprachigen YouTube-Kommentaren des NottDeuYTSch-Korpus (Cotgrove 2023) manuell annotiert. Etwa ein Drittel der Kommentare in diesem Datensatz enthält Flauschäußerungen. Das legt nahe, dass Flausch einen relevanten Anteil der Onlinekommunikation ausmacht und zur Attraktivität und Beliebtheit digitaler Plattformen beitragen kann.

Keywords: digitale Kommunikation, YouTube, Flausch, unterstützende Sprachhandlungen, Deutsch, Korpusdaten, Annotation

Kontaktpersonen:
Yulia Clausen & Tatjana Scheffler
Ruhr-Universität Bochum
Universitätsstr. 150
44780 Bochum
yulia.clausen@rub.de
tatjana.scheffler@rub.de

ISSN: 2569-6491
CC BY-SA 4.0

1 Einleitung

Soziale Medien stellen einen festen Bestandteil des Alltags dar. Ihre weitreichende gesellschaftliche Relevanz lässt sich unter anderem darauf zurückführen, dass sie den Zugang zu virtuellen Lebenswelten eröffnen (Rieger/Schäfer/Tuschling 2021). In solchen Kontexten finden Nutzer*innen häufig die Freiheit, sich neu bzw. anders zu gestalten (Kersten/Lotze 2019) und dadurch unterschiedliche (virtuelle) Persönlichkeiten auszuleben (Burnett 2017; 2019). So können sie an verschiedenen virtuellen Gemeinschaften (oder Communities; Rheingold 1994) teilhaben bzw. neue gründen und sich auf diese Weise vielfältige virtuelle Lebenswelten erschließen. Diese stehen jedoch nicht im Widerspruch zu physischen Welten, sondern erweitern sie, insbesondere durch die Pluralität der Kontexte und die sozialen Gemeinschaften, in denen sie verortet sind (vgl. Chayko 2020; Leuckert 2020). Virtuelle Gemeinschaften verstehen wir als solche, die sich unabhängig von örtlichen und zeitlichen Beschränkungen durch Gruppenbildungsprozesse und soziale Interaktion herausbilden, zum Beispiel in Form von Interessensgruppen oder Fan-Communities.

Virtuelle Phänomene sind dabei eng mit dem Alltagsleben verflochten. Das hat zur Folge, dass ein immer größerer Anteil der sprachlichen Kommunikation digital stattfindet und nicht von Angesicht zu Angesicht im physischen Raum.[1] Unser Forschungskontext beginnt mit der Beobachtung, dass Individuen täglich viel Zeit in sozialen Medien verbringen[2] und große Mengen sprachlicher und nicht-sprachlicher Interaktionen im Internet stattfinden. Dabei gilt digital vermittelte sprachliche Kommunikation oft als weniger stark sozial eingebettet im Vergleich zu den nicht-digitalen Kommunikationsformen. Diese Kommunikation ist im Unterschied zu Face-to-Face Kommunikation größtenteils einseitig, da Creator*innen auf Plattformen wie YouTube oder Instagram typischerweise nicht auf Kommentare reagieren, und einzelne Kommentare zudem selten in Diskussionen eingebunden werden (Merten 2022). Dies hat zur Folge, dass an sozialen Normen orientierte Verhaltensweisen in Online-Kontexten oft weniger ausgeprägt sind als in der Face-to-Face Kommunikation, wodurch die Hemmschwelle für negative Interaktionen und Hassrede sinken kann (*online disinhibition effect*; Suler 2004). In virtuellen Interaktionen verhalten sich Menschen daher

1 In Deutschland nutzen rund 90% der Personen ab 16 Jahren regelmäßig Online-Kommunikationsdienste (Stand 2021; Bundesnetzagentur, 2023).

2 Im Schnitt verbringen Nutzer*innen weltweit täglich fast 2,5 Stunden auf Social-Media-Plattformen (siehe https://www.statista.com/statistics/270229/usage-duration-of-social-networks-by-country/; Stand 2023; abgerufen am 22.05.2025).

teils direkter oder auch aggressiver, da sie sich anonymer fühlen und weniger unmittelbare Konsequenzen erwarten (Jaki 2023). Dabei sind soziale Medien keine reinen Orte sozialer Vereinzelung, an denen Nutzer*innen sich Hass und Hetze aussetzen. Individuen verbringen dort regelmäßig Zeit, um positive Inhalte sowie ihre Lebenswelten zu teilen (siehe z. B. Biri 2023 zu affektiven Positionierungen in Online Fan-Communities).

Wir fokussieren uns auf positive Onlinesprache, insbesondere auf einen unterstützenden und gruppenkonstituierenden Austausch, den wir als „Flauschrede“ (*candy speech*; Clausen/Scheffler 2025) bezeichnen. Flausch umfasst – wie im Folgenden gezeigt wird – mehrere Untertypen, darunter Komplimente, Zuneigungsbekundungen und Dankbarkeitsausdrücke. Solche Flauschtypen sind mit Höflichkeitskonzepten eng verknüpft und beinhalten daher teils auch konventionalisierte, formelhafte Ausdrucksweisen und Wendungen (Merten 2022), wie z. B. die Äußerungen in (1–3) in dem von uns verwendeten Korpus.[3] Im Vordergrund steht bei Flausch jedoch die durch diese Ausdrücke realisierte Unterstützungsinteraktion, sodass Höflichkeit dabei keine zentrale Rolle spielt (seihe aber z. B. Graham 2007; Darics 2010 zu (Un)höflichkeit in der Onlinekommunikation).

(1) danke das du dir so viel Mühe machst ❤
(2) du bist schön ;*
(3) xoxoxo

(NDY/138)

Wie unsere Korpusdaten zeigen, handelt es sich bei Flauschrede um einen zentralen und allgegenwärtigen Aspekt virtueller Kommunikation (siehe Abschnitt 4), die allerdings bisher eine deutlich geringere Wahrnehmung und Aufmerksamkeit erhalten hat als negative Formen der Interaktion wie Hassrede. Zu Hassrede und ihrer (automatischen) Erkennung liegt eine Vielzahl an Studien vor (siehe u. a. Hosseinmardi et al. 2015; Brodnig 2016; Schmidt/Wiegand 2017; Wiegand et al. 2019; Ashraf et al. 2021; Kapil/Ekbal 2024). Positive, unterstützende Onlinesprache ist hingegen in der (Computer-)Linguistik noch weitgehend unerforscht (für Ausnahmen siehe Abschnitt 3.2). Auch in der formalen Linguistik liegen Untersuchungen

3 Alle Beispiele in dieser Studie entstammen dem in Abschnitt 4 vorgestellten NottDeuYTsch-Korpus. Die Kommentare wurden originalgetreu inklusive aller Rechtschreibfehler übernommen. Das Sigle (NDY/xxx) zeigt, dass es sich um einen Kommentar zum Video xxx aus diesem Korpus handelt.

zur Wirkungsweise bzw. Modellierung von Expressivität bislang primär im Kontext negativer Emotionalität vor (siehe Scheffler 2025). Ein zentrales Anliegen der vorliegenden Studie ist es, dieses Ungleichgewicht zu adressieren.

Im Hinblick auf Medialität stellt sich die Frage, inwiefern sich unterschiedliche Plattformen (z. B. Instagram vs. YouTube) hinsichtlich der sprachlichen Konstruktion von Communities sowie des Auftretens und der Verwendung positiver Sprache unterscheiden. Grundsätzlich lassen sich sowohl community-spezifische, als auch Plattform-übergreifende stilistische Merkmale identifizieren, die zum Beispiel den Schwierigkeitsgrad der Beiträge (höher in Reddit als 4chan) oder die Verbreitung von Schimpfwörtern (höher in 4chan als Reddit) betreffen (Khalid/Srinivasan 2020). Daraus lässt sich ableiten, dass positive Kommunikation in bestimmten Medien verbreiteter ist als in anderen und sich die Plattformen auch durch die Art, in der sprachliche Gruppenbildung stattfindet (z. B. primär durch gemeinsame Beleidigung Außenstehender oder durch gegenseitige Unterstützung), unterscheiden. Eine systematische Untersuchung dieser Unterschiede wird bislang dadurch erschwert, dass es an einer differenzierten Definition positiver Onlinesprache mangelt und bisher nur grobe Sentiment-Analysen als Annäherung verfügbar sind. Die vorliegende Studie setzt hier an, um die Grundlage für zukünftige Medium-vergleichende Analysen zu schaffen.

Im Folgenden erarbeiten wir, basierend auf der Etymologie des Begriffs „Flausch“, eine Definition dieses Phänomens sowie seiner Untertypen. Wir leiten Erkennungsmerkmale ab und untersuchen das Vorkommen sowie die Verwendung von Flauschrede in einem Korpus deutscher YouTube-Sprache (NottDeuYTsch; Cotgrove 2023). Das Korpus umfasst einen Ausschnitt internetbasierter Kommunikation aus den Jahren 2008 bis 2018, ist jedoch weiterhin aussagekräftig für die vorliegende Untersuchung. Unser Ziel ist es, die Relevanz von Flauschrede als virtuellem sozialem Phänomen hervorzuheben und ein Bewusstsein für ihre Präsenz in sozialen Medien zu schaffen. Das kann – neben der Eindämmung von Hassrede – eine ergänzende Strategie sein, um zur positiven Gestaltung virtueller Kommunikation beizutragen. Konkret geht unsere Studie der Aufgabe nach, eine Basis für die Annotation und automatische Erkennung von Flauschrede zu schaffen.

2 Zum Begriff „Flausch“

Der Begriff „Flausch“ stammt aus dem 18. Jahrhundert und bezeichnete ursprünglich einen „weichen, haarigen Wollstoff“.[4] Das Wort leitet sich von *vlūs*, *vlūsch* („wolliges Schaffell“) ab und bezieht sich auf ein dickes, weiches, stark gerautes Wollgewebe, das für Mäntel verwendet wurde. Parallel dazu wird das Adjektiv „flauschig“ („weich, wollig“; 19. Jahrhundert) im wörtlichen Sinne für weiche, kuschelige Stoffe und Plüschtiere sowie auch für kleine Tiere verwendet.[5] Das Substantiv ist heute im ursprünglichen Sinn eher ungebräuchlich und leitet sich in seiner Bedeutung vom Adjektiv ab.

Der Begriff „Flausch“ wurde in digitalen Medien spätestens ab 2009 durch Stefan Niggemeiers Blogrubrik „Flausch am Sonntag“ bekannt.[6] Beiträge, die 2008 zunächst als „Katzencontent“[7] veröffentlicht wurden, haben sich im Laufe der Zeit zum sogenannten „Flauschcontent“ entwickelt. Dabei handelt es sich um Video- und Fotobeiträge, die (im wörtlichen Sinne) flauschige Tiere zeigen, seien diese real, gezeichnet oder aus Plüsch.

Spätestens im Jahr 2012 wurde der Begriff von der Piratenpartei aufgegriffen. In der Zwischenzeit hatte er eine deutliche Bedeutungserweiterung erfahren und wurde nicht länger ausschließlich im Wortsinne als Bezeichnung für Tiere mit weichem Fell verwendet, sondern auch metaphorisch auf Inhalte übertragen, die allgemein als süß, positiv oder emotional wohltuend empfunden werden. In der Terminologie der Piratenpartei wurde Flausch im Sinne eines „erträglichen und netten Miteinanders gemäß der Internet-Netiquette“ definiert und mit Konzepten wie Fürsorge, Entspannung und „MITeinander statt GEGENeinander“ assoziiert. In diesem Kontext gründete die Partei die „AG Flausch“, deren Ziel es ist, „Mittel und Wege [zu] entwickeln um für einen flauschigeren Umgang unter den Mitgliedern der Piratenpartei zu sorgen“.[8] In Anlehnung an diese Idee bietet die AG eine „Portion Flausch“ an, sollte es mal „Not am Manne oder an der Frau“ sein.

4 „Flausch“, bereitgestellt durch das Digitale Wörterbuch der deutschen Sprache, https://www.dwds.de/wb/Flausch (abgerufen am 12.03.2025).

5 „flauschig“, in: Wolfgang Pfeifer et al., Etymologisches Wörterbuch des Deutschen (1993), https://www.dwds.de/wb/etymwb/flauschig (abgerufen am 12.03.2025).

6 Der erste Flausch-Post: https://www.stefan-niggemeier.de/blog/5828/flausch-am-sonntag-1/. Siehe auch: https://taz.de/Der-grosse-Flauschangriff/!578329/.

7 https://www.stefan-niggemeier.de/blog/721/katzencontent/.

8 https://wiki.piratenpartei.de/AG_Flausch (abgerufen am 12.03.2025).

Auch als Hashtag wird Flausch in verschiedenen Wortbildungen, wie z. B. *#FlauschTheCurve*[9] verwendet. Damit ist der Begriff „Flausch“ im allgemeinen Sprachgebrauch der Internetgemeinschaft angekommen und bezeichnet generell unterstützende und positive Beiträge in sozialen Medien und im Netz. Beispielhaft kann hierfür ein Beleg aus dem von uns verwendeten Korpus aus dem Jahr 2017 herangezogen werden (4).

(4) Geiles video #Glocke aktiv #daumen nach oben 👍 #hdl 💖
Fände es flauschig wenn du mich grüßen könntest 💋 ❤️
(NDY/202)

3 Definition von Flausch und Flauschtypen

Ausgehend von der intuitiven Nutzung des Begriffs „Flausch“ für unterstützende Online-Beiträge erarbeiten wir eine Definition von Flausch als linguistischer Kategorie, insbesondere in Abgrenzung zu anderen Formen positiver Onlinesprache, und beschreiben dessen verschiedene Typen.

3.1 Definition

Wir definieren Flausch als Ausdruck positiver Einstellungen und Gefühle in sozialen Medien gegenüber Einzelpersonen, d. h. Contentcreator*innen (in unseren Daten Autor*innen von YouTube Videos) und Kommentator*innen oder deren Beiträgen (Videos und Kommentare). Der Zweck von Flausch besteht darin, andere Personen zu ermutigen, aufzumuntern, zu unterstützen oder zu bestärken (siehe auch Clausen/Scheffler 2025). Face Work bezeichnet (sprachliche) Handlungen, die das soziale Selbstbild von Sprecher*innen oder Adressat*innen positiv oder negativ beeinflussen. Demnach zielt Flausch darauf ab, das Selbstbild (*Face*; Goffman 1967) der Zielperson/-gruppe auf positive Weise zu beeinflussen.

Analog zur toxischen Sprache und zur Hassrede, die häufig durch ihre Effekte (z. B. Einschüchterung, Ausgrenzung oder emotionale Verletzung) sowie durch die Intentionen ihrer Verwender*innen (z. B. Provokation oder Abwertung) charakterisiert werden (Jaki 2023), betrachten wir den intendierten perlokutiven Effekt des Flauschs, nämlich die Bestärkung des positiven Selbstbildes einer Person oder Gruppe, als definierendes Kriterium dieses Begriffs.

9 Siehe z. B. https://tschop-tschop.de/flausch-gegen-die-angst/index.html (abgerufen am 05.04.2026).

Flausch lässt sich demnach als kommunikative Handlung verstehen, die unterstützendes *Face Work* im Sinne Goffmans leistet.

3.2 Abgrenzung von ähnlichen Konzepten

Mehrere Begriffe werden im Zusammenhang mit positiver Sprache und phatischer Kommunikation verwendet. Wir konzentrieren uns hier auf Kategorisierungen von Sprachhandlungen, die auf Online-Interaktionen anwendbar sind.

Der Interaktionstyp Smalltalk (oder *social talk*) fällt in den Bereich der phatischen Kommunikation und dient weniger dem Informationsaustausch als der „Herstellung von sozialen Kontakten“ (Klüwer 2011: 14). Klüwer (2011) schlägt eine Taxonomie von Dialogakten vor, die auf kooperative soziale Gespräche in virtuellen Welten (nach Art von *Second Life*) ausgelegt ist. Diese Taxonomie basiert auf der Theorie des Face Work – auf die wir ebenfalls als theoretische Grundlage zurückgreifen – sowie auf Konzepten unterstützender Dialogakte. Da sowohl Smalltalk als auch Flausch das Ziel haben, Beziehungen zu pflegen und damit das „Gesicht“ des Gegenübers zu wahren, übernehmen wir einige für unsere Analyse relevante Kategorien aus Klüwer (siehe Abschnitt 3.3), die als Ausgangspunkt für unsere Taxonomie dienen. Die beiden Formen unterscheiden sich jedoch insofern, als Flausch in der Regel deutlich stärkere emotionale Unterstützung bietet (siehe Abschnitt 3.1). Smalltalk beinhaltet hingegen beispielsweise sprachliche Akte, die Interesse am Gesprächspartner signalisieren, indem gezielte Nachfragen gestellt werden, ohne dabei eine direkte emotionale Unterstützung zu leisten. Solche Beiträge würden nicht unter Flauschrede fallen, da ein wesentlicher perlokutiver Effekt in der inhaltlichen Auseinandersetzung liegt.

Hinsichtlich der Bedeutung und Verwendung lässt sich Flausch mit dem Begriff „Candystorm“ vergleichen:

> Mit dem Candystorm geht eine Welle des Zuspruchs im virtuellen Raum einher, z. B. in sozialen Netzwerken, Microblogs und Blogs sowie Kommentarbereichen von Onlinezeitungen und -zeitschriften. Er wird evoziert durch den Moralismus der Informationsgesellschaft und die Empathie und Euphorie der Netzbürgerinnen und -bürger. Personen oder Organisationen werden mit Worten des Zuspruchs und Begriffen wie „Flausch“ bedacht. Das Gegenteil ist der Shitstorm (Bendel 2019: 37).

Wie aus dieser Definition hervorgeht, bezeichnet Candystorm wellenartige Ausbrüche entsprechender sprachlicher Dynamiken. Der

Begriff „Flauschrede“ ist dagegen neutraler gefasst und lässt sich auch auf kleinere Einheiten unterstützender Kommunikation anwenden, beispielsweise auf einzelne Kommentare.

Eine Form phatischer Onlinesprache stellt „bestärkende Sprache“ (*empowering language;* Njoo et al. 2023) dar. Diese Sprache hat zum Ziel, „die Rechte, die Wahlmöglichkeiten, die Selbstverwirklichung oder das Selbstwertgefühl von Adressat*innen zu fördern“ (Njoo et al. 2023: 9335; unsere Übersetzung). Ähnlich wie beim Flausch handelt es sich dabei um unterstützende Sprachhandlungen, die das positive Selbstbild der Adressat*innen stärken. Da der Fokus hierbei jedoch ausschließlich auf der Stärkung der gefühlten Handlungsmacht der Adressat*innen liegt, ergibt sich nur eine partielle Überschneidung mit unserer Definition von Flausch. So werden in der zitierten Studie Kommentare wie *And know that it's ok to be sad. It's ok to mourn the loss of someone you love - whether they were great or they were awful. Its normal and you shouldn't feel weird or bad because of it. Good luck :)* als bestärkend klassifiziert (Njoo et al. 2023: 9334). Obwohl solche Beiträge darauf abzielen, bei den Adressat*innen ein positives Gefühl hervorzurufen, richten sie sich nur indirekt auf die Unterstützung ihres sozialen Selbstbildes. Viele der in der Studie als bestärkend klassifizierten Aussagen würden daher nicht notwendigerweise unter Flausch im Sinne unserer Definition fallen.

Eine weitere Form positiver Onlinekommunikation, die insbesondere im Bereich der automatischen Erkennung und Klassifikation Aufmerksamkeit erhält, ist die Hoffnungsrede (*hope speech*). Sie bezeichnet eine Sprachform, die „positiv, ermutigend, beruhigend, integrativ und unterstützend ist und Menschen inspiriert sowie Optimismus vermittelt“ (Chakravarthi et al. 2022: 379; unsere Übersetzung). Diese Definition zeigt, dass Hoffnungsrede in bestimmten Aspekten mit dem Konzept des Flauschs übereinstimmt, jedoch einen inhaltlich abweichenden Schwerpunkt setzt. Sie bezieht sich ausschließlich auf sprachliche Äußerungen, die der Entschärfung von Feindseligkeiten dienen (vgl. Chakravarthi 2020; Palakodety et al. 2020; García-Baena et al. 2023). Im Gegensatz dazu setzt Flausch nicht zwingend eine vorangehende feindliche oder negative Kommunikation voraus, sondern kann auch unabhängig davon in unterstützender Absicht eingesetzt werden.

Darüber hinaus bilden Komplimente als eine Form emotionaler Unterstützungspraktiken einen in der Forschung intensiver untersuchten Gegenstand (siehe u. a. Maíz-Arévalo/García-Gómez 2013; Placencia/Lower 2017; Placencia/Eslami 2020; Merten 2022). Die einschlägigen Studien operieren jedoch mit unterschiedlichen und teilweise sehr weit gefassten Definitionen des Kompliments, die ein breites Spektrum an Ausdrucksformen umfassen – von Emojis und

minimalen Rückmeldesignalen wie *Mmmh* bis hin zu verschiedenen Typen positiver Bewertungen (siehe Placencia/Lower 2017: 635–637 für eine Diskussion). In unserer Studie werden Komplimente als ein Untertyp von Flauschrede konzeptualisiert, der anhand spezifischer, enger gefasster Kriterien definiert wird (siehe Abschnitt 3.3). Vor diesem Hintergrund ist die Vergleichbarkeit mit den genannten Arbeiten nur eingeschränkt gegeben, da dort als Komplimente eingeordnete Ausdrucksformen im vorliegenden Ansatz auf mehrere distinkte Untertypen von Flauschrede verteilt sind.

3.3 Typen von Flausch

Wir betrachten Flauschkommunikation als sprachliche Handlungen mit einer bestimmten Intention, die sich daher in Subtypen unterteilen lassen. Für die feingliedrige Klassifikation von Flauschausdrücken greifen wir, wie bereits erwähnt, auf die Taxonomie der Smalltalk-Dialogakte nach Klüwer (2011) zurück. Klüwer berücksichtigt in ihrer Analyse Formen des Face Work, die sowohl auf das eigene Gesicht der Sprecher*innen abzielen, als auch solche, die auf das Gesicht der Adressat*innen gerichtet sind. Im Fall von Flausch ist jedoch vor allem das adressatenbezogene Face Work relevant. Vor diesem Hintergrund übernehmen wir aus Klüwers Taxonomie jene Kategorien, die mit unserer Definition von Flausch übereinstimmen, wie zum Beispiel Komplimente und positives Feedback, und erweitern sie um zusätzliche Typen, die auf der Grundlage unserer Korpusanalyse identifiziert wurden. Im Gegensatz zu Klüwer annotieren wir in der empirischen Anwendung keine ganzen Dialogakte, sondern ausschließlich die darin enthaltenen Flauschausdrücke. Unsere Klassifikation umfasst zehn Kategorien: *affection declaration*, *compliment*, *encouragement*, *gratitude*, *group membership*, *positive feedback*, *sympathy*, *implicit*, *agreement* und *ambiguous*. Die einzelnen Flauschtypen sind so definiert, dass sie sich gegenseitig ausschließen und eine eindeutige Klassifikation ermöglichen. So beziehen sich Komplimente ausschließlich auf die positive Bewertung persönlicher Eigenschaften und nicht auf Beiträge (z. B. Videos), Kleidungsstücke oder andere im Video gezeigte Gegenstände. Letzteres wird durch *positive feedback* erfasst, das als einzige Kategorie allgemeiner Natur gilt und potenziell weiter differenziert werden könnte, ohne sich mit den übrigen Flauschtypen zu überschneiden.[10]

10 Eine solche Differenzierung könnte beispielsweise in Bezug auf formale Unterschiede (linguistische Ausdrücke vs. alleinstehende Emojis) oder Unterschiede auf der syntaktischen Ebene erfolgen, etwa zwischen Ich-Formulierungen wie „Ich finde es cool“ und Äußerungen wie „Mega süß!!!“, die kein explizites Bewertungssubjekt enthalten. Daran ließen sich weitere pragmatische und medien-linguistische Analysen anknüpfen (vgl. Merten 2022).

Im Folgenden werden die einzelnen Flauschtypen definiert und anhand von Beispielen veranschaulicht. Vollständige Definitionen und Erläuterungen befinden sich in unseren ausführlichen Annotationsrichtlinien.[11]

Affection declaration – Bewunderungs-, Liebes- und Zuneigungs-Bekundungen gegenüber Personen, d. h. Contentcreator*innen oder Kommentator*innen (nicht aber gegenüber deren Beiträgen, wie Videos oder Kommentaren). Diese Äußerungen müssen nicht notwendigerweise romantisch gemeint sein.

(5) ich liebe dich so sehr würde mich sehr freuen wen du mir antworten könntest und mich grüsenn könntest #beste youtuberin[12]

(NDY/202)

Compliment – Komplimente und freundliche Bemerkungen, die sich auf andere Personen und nicht auf deren Beiträge beziehen. Komplimente können Fähigkeiten (z. B. Singen, Kochen), persönliche Eigenschaften (z. B. Stimme, Aussehen) oder Leistungen (allgemein oder im Kontext eines bestimmten Videoauftritts) sowie deren Bewertung betreffen.

(6) deine stimme ist sehr angenehm . du bist wunderschön !

(NDY/020)

Encouragement – Äußerungen, die darauf abzielen, andere Personen (Contentcreator*innen und Kommentator*innen) zu ermutigen. Dies kann sich auf das Weitermachen in sozialen Medien (z. B. weiteres Posten oder Liken von Beiträgen), auf die Stärkung von Vertrauen oder auf die Verringerung bzw. Beseitigung von Sorgen, Unsicherheit und ähnlichen Gefühlen beziehen.

(7) Ich finde das ist sehr Geil weil am Anfang ist das sehr gut also macht immer so weiter

(NDY/179)

11 Die Annotationsrichtlinien sind via OSF verfügbar: https://osf.io/efh7c/.

12 In Kommentaren, die über die eigentlichen Flauschausdrücke hinausgehenden Text enthalten, werden die entsprechenden Flauschausdrücke durch Unterstreichung hervorgehoben.

Gratitude – Ausdrücke aufrichtiger Dankbarkeit, die unaufgefordert bzw. spontan geäußert werden. Diese Dankesäußerungen sind deutlich abzugrenzen von routinierter Höflichkeit, wie reaktivem Danken (z. B. „Danke für die Antwort“) oder formelhaften Ausdrücken wie „Danke im Voraus“.

(8) Danke , dass du das für uns machst 😍 💖 💞

(NDY/033)

Group membership – Ausdrücke, mit denen eine bestimmte Gruppenzugehörigkeit signalisiert wird, wie zum Beispiel zur Fangemeinschaft eines/r bestimmten Contentcreator*in. Dazu zählen insbesondere Hashtags und Emojis, die als Markenzeichen der jeweiligen Creator*innen oder als Erkennungszeichen der jeweiligen Community fungieren, aber auch explizite, wörtliche Bekundungen der Zugehörigkeit.[13]

(9) Einhörner 🦄 #BonnyTrash[14]

(NDY/202)

Positive feedback – Äußerungen, die eine positive Haltung gegenüber dem ursprünglichen Beitrag oder einem Kommentar ausdrücken, jedoch nicht gegenüber einer spezifischen Person. Auch Emojis und Emoticons, die ohne begleitende sprachliche Äußerungen gepostet werden, fallen in diese Kategorie.

(10) Könnt ihr mal vielleicht wieder ein fakten video über euch machen wäre mega geil und die parodie ist mega hammer geil 😍 😍 😍 😍 😍 😍

(NDY/186)

Sympathy – Äußerungen, die Mitgefühl und Verständnis ausdrücken. Diese können als Reaktionen auf negative oder betrübliche

13 Wir verstehen Gruppenzugehörigkeit im engeren Sinne, wobei sie gemäß unserer Definition durch etablierte Gruppenkennzeichen oder explizite Fan-Deklarationen (die für die jeweilige Community charakteristisch sind) signalisiert wird.

14 Zum Einhorn als Gruppenkennzeichen der BonnyTrash-Fangemeinde, siehe auch Cotgrove (2025, Kapitel 6).

Mitteilungen anderer Personen erfolgen, aber auch initiativ Sympathie ausdrücken, wie zum Beispiel in Bezug auf eine bestimmte Lage oder Situation, die die Zielperson betrifft.

(11) + Elma D. Oh nein ! 🙈 Aufwärmen ist so so wichtig !!! Ich hoffe es geht dir langsam besser !

(NDY/033)

Die folgenden drei Annotationskategorien beziehen sich auf Fälle, die zwar nicht unmittelbar Flausch ausdrücken, jedoch für die Analyse als relevant erachtet und daher mit annotiert wurden.

Implicit – Flausch wird nicht direkt geäußert, sondern kann aus einem bestimmten wörtlichen Ausdruck abgeleitet werden. Damit ein Ausdruck als implizite Form von Flausch annotiert werden kann, muss eindeutig erkennbar sein, welchem der zuvor definierten Flauschtypen er zuzuordnen ist. Auf diese Weise erfasst diese Kategorie alle Flauschtypen und könnte feingliedriger annotiert werden, indem der jeweils vorliegende Flauschtyp (z. B. *implicit-compliment*) angegeben wird, der jedoch nur implizit zum Ausdruck kommt.[15]

(12) Wieso macht ihr nicht bei dsds oder bein supertalent mit ? ihr seit sooooo gut

(NDY/003)

Agreement – Ausdruck von Zustimmung zu der Meinung oder Aussage einer anderen Person. Dabei werden ausschließlich Kommentare berücksichtigt, die sich nachweislich auf Flauschäußerungen beziehen. Diese Kategorie signalisiert eine positive Reaktion auf den von anderen verfassten Flausch, sodass anzunehmen ist, dass die zustimmende Person den Flausch, dem sie zustimmt, potenziell auch selbst hätte verfassen können.

(13) Bin ganz deiner Meinung :)

(NDY/179)

Ambiguous – diese Kategorie wurde in den Fällen verwendet, in denen nicht eindeutig bestimmt werden konnte, ob es sich um eine

15 Dass *implicit* per Definition andere Flauschtypen unserer Taxonomie umfasst, betrachten wir nicht als definitorische Überschneidung, sondern als Folge der von uns gewählten Kategorisierung.

Flauschäußerung handelt, zum Beispiel aufgrund fehlenden Kontexts oder mehrdeutiger Ausdrucksweise. Wir dokumentieren diesen Fall, wobei der weitere Umgang damit von den Zielen der jeweiligen Analyse abhängt.

(14) OMG

(NDY/247)

Ergänzend zu den definierten Flauschtypen wurde das Label *uncertain* eingeführt, um Unsicherheiten und potenziell mehrdeutige Fälle zu kennzeichnen, die anschließend besprochen und geklärt wurden.

4 Empirische Analyse

4.1 Datengrundlage

Als Datengrundlage für die vorliegende Untersuchung dient das NottDeuYTSch-Korpus (Cotgrove 2023). Das Korpus umfasst über 33 Millionen Wörter aus rund 3 Millionen YouTube-Kommentaren, die im Zeitraum von 2008 bis 2018 von jungen, deutschsprachigen Nutzer*innen verfasst wurden. Für unsere Analyse haben wir einen Ausschnitt der Daten bestehend aus Kommentaren zu insgesamt 16 Videos von sieben verschiedenen Contentcreator*innen ausgewählt (siehe Tabelle 1).[16] Die Metadaten der einzelnen Dokumente umfassen unter anderem Angaben wie den Kanalnamen bzw. den Namen des Contentcreators/der Contentcreatorin, den Videotitel, das Veröffentlichungsdatum des Videos, die Namen der Kommentierenden sowie das Datum der Kommentare. Um eine diverse Datenbasis zu erzielen, wurden Dokumente von YouTube-Kanälen unterschiedlicher Thematiken ausgewählt, sodass davon ausgegangen werden kann, dass die Kommentierenden verschiedene Communities repräsentieren. Bei der Erstellung der Datenbasis haben wir zudem auf ein ausgewogenes Verhältnis zwischen einer breiten Anzahl von erfassten Creator*innen und der Möglichkeit der Detailanalyse einzelner Creator*innen (in Form mehrerer Videos bzw. deren Kommentare) geachtet. Unser Datensatz lässt sich darüber hinaus auf beiden Ebenen erweitern.

16 Die Daten sind über einen Download verfügbar: http://hdl.handle.net/11372/LRT-4779.

Creator*in	Video-Themen	Datei	Kommentare gesamt / davon mit Flausch	Flausch-ausdrücke
Anna Maria Damm	Lifestyle Beauty Fashion Pflege-Routinen	NDY/285	203 / 128	184
DieLochis	Musik Songparodien Unterhaltung	NDY/003 NDY/004 NDY/179 NDY/186 NDY/208	32 463 / 8 607	11 204
Nilam	Beauty Lifestyle Alltags- und Kreativ-Vlogs	NDY/020 NDY/138 NDY/250 NDY/252 NDY/274	8 385 / 3 806	6 636
SophiaThiel	Fitness Bodybuilding Ernährung	NDY/033	2 387 / 813	1 310
xKarenina	Schmink-Tutorials Buch- und Produkt-Vorstellungen	NDY/246	241 / 86	117
BonnyTrash	Lifestyle Vlogs Challenges	NDY/202 NDY/203	925 / 638	1 523
Sami Slimani	Mode Styling Beauty Pflege	NDY/247	1 682 / 502	811
Gesamt			46 286 / 14 580	21 785

Tabelle 1: Übersicht der Datengrundlage

Die zu den ausgewählten Videos geposteten Kommentare wurden hinsichtlich des Vorkommens von Flausch analysiert und gemäß der in Abschnitt 3.3 erläuterten Flauschtypen manuell mittels des Annotationstools INCEpTION (Klie et al. 2018) annotiert.[17] Die Annotationen wurden von zwei Annotator*innen durchgeführt (Annotatorin

17 Der annotierte Datensatz steht für Forschungszwecke via OSF frei zur Verfügung: https://osf.io/efh7c/.

I, eine Autorin dieses Beitrags, und Annotator II, eine Person mit linguistischem Hintergrund), wobei Annotator II den Großteil der Daten annotierte. In der initialen Einarbeitungsphase annotierten die beiden Annotator*innen unabhängig ein Dokument, und die Ergebnisse wurden anschließend gemeinsam besprochen. Im weiteren Verlauf des Annotationsprozesses wurden die Annotationen des Annotators II in regelmäßigen Abständen von Annotatorin I überprüft, und alle festgestellten Inkonsistenzen wurden gemeinsam diskutiert und geklärt. Neue Erkenntnisse und hinzugekommene Kategorien wurden in die bereits annotierten Daten rückwirkend integriert. Zur Bewertung der Übereinstimmung zwischen den Annotator*innen (*inter-annotator agreement*) wurde für die binäre Entscheidung, ob ein Kommentar Flausch enthält, der Cohen's Kappa Wert berechnet. Die Berechnung erfolgte auf Grundlage von zwei Dokumenten, die nicht Teil der Einarbeitungsphase waren und von den Annotator*innen unabhängig voneinander annotiert wurden. Die ermittelten Kappa-Werte betragen 0,7 bzw. 0,76, was auf eine hohe Übereinstimmung zwischen den Annotator*innen hinweist (für Details siehe Clausen/Scheffler 2025).

Die annotierten Daten liefern Einblicke in die Verteilung der Flauschtypen und sollen sowohl das Bewusstsein für deren Vorkommen im virtuellen Diskurs fördern als auch eine Grundlage für computerlinguistische Ansätze bilden. Beispielsweise haben wir den Datensatz (als Trainings- und Testdaten) sowie die Annotationsrichtlinien im Rahmen einer Shared Task zur Verfügung gestellt (Clausen/Scheffler/Wiegand 2025), um die automatische Erkennung und Klassifizierung dieser Art positiver Interaktionen in sozialen Medien zu erforschen. Im folgenden Abschnitt wird die Verteilung der annotierten Flauschtypen in den Daten näher beschrieben und diskutiert.

4.2 Verteilung der Flauschtypen

Es wurden 46 286 Kommentare hinsichtlich Flauschäußerungen analysiert. In 14 580 dieser Kommentare konnte mindestens ein Flauschausdruck identifiziert werden, so dass insgesamt 21 785 Flauschausdrücke annotiert wurden. Abbildung 1 zeigt die relative Häufigkeitsverteilung der Flauschtypen in den Daten.

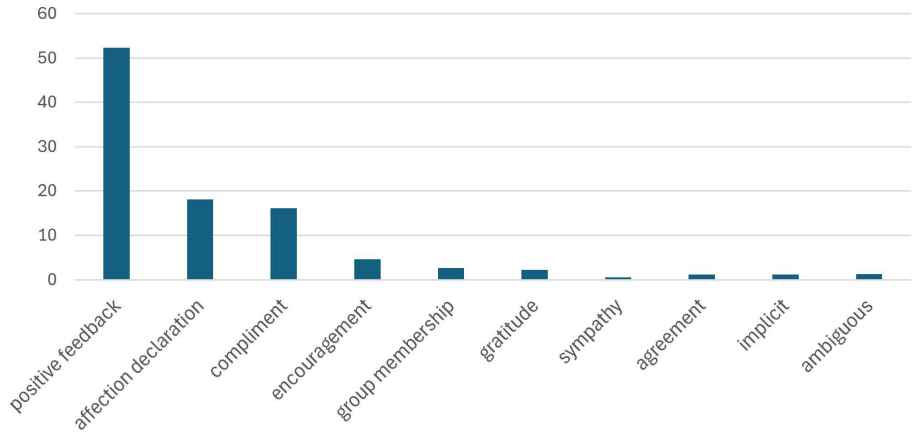


Abbildung 1: Gesamtverteilung der Flauschtypen (in %)

Positive feedback ist mit 52,3% der am häufigsten vertretene Flauschtyp und macht damit etwas mehr als die Hälfte aller annotierten Flauschausdrücke aus. Das liegt unter anderem daran, dass diese Kategorie Äußerungen „allgemeiner Art" umfasst, beispielsweise solche, die positive Gefühle oder Einstellungen ausdrücken, ohne ein spezifisches Bezugsobjekt (Person oder Beitrag) zu benennen, wie zum Beispiel *voll geil* oder *cool*. Ebenfalls in diese Kategorie fallen Emoticons und Emojis, die häufig eigenständig – also ohne begleitende sprachliche Elemente – gepostet werden. Diese Form nonverbaler Flauschrede lässt sich mit geringem Aufwand einsetzen und tritt daher in der untersuchten YouTube-Kommunikation besonders häufig auf. Emoticons und Emojis mit positiver Konnotation gelten als unterstützendes Face-Work („gesichtsbestärkende Mittel"; Beißwenger/Pappert 2019) und können daher als Flausch klassifiziert werden. In unseren Daten machen sie rund 2,4% der Äußerungen innerhalb der Kategorie *positive feedback* aus.

Die zweithäufigsten Flauschtypen sind solche, die sich auf Personen beziehen, beispielsweise durch positive Bewertungen persönlicher Eigenschaften und Leistungen (*compliment*; 16,1%) oder durch direkte Ausdrücke der Bewunderung und Zuneigung (*affection declaration*; 18,1%). Zusammengenommen machen diese personenbezogenen Formen des Flauschs etwa 34% aller annotierten Äußerungen aus und kommen somit seltener vor als allgemeinere positive Rückmeldungen.

Die übrigen Flauschtypen bilden jeweils weniger als fünf Prozent aller Flauschäußerungen: *encouragement* (4,6%), *group membership* (2,6%), *gratitude* (2,2%) und *sympathy* (0,5%). Die Kategorie *encouragement* dient primär dazu, YouTuber*innen zum Durchhalten oder Weitermachen zu motivieren. Obwohl dieser Flauschtyp se-

mantisch relativ allgemein gestaltet ist, tritt er seltener auf als die zuvor beschriebenen Kategorien.[18] Insbesondere findet er sich in Kommentaren zu Videos, deren Inhalte diese Art unterstützender Botschaften begünstigen, beispielsweise bei angehenden Sänger*innen. Auch Kommentator*innen können adressiert und zu positiver Interaktion, wie zum Beispiel dem Geben von Likes an den/die Creator*in, ermutigt werden. Die anderen Flauschtypen sind inhaltlich enger gefasst. So signalisiert *group membership* die Zugehörigkeit zu bestimmten Communities, wie Fangruppen oder thematisch-orientierten Interessengemeinschaften. In unseren Korpusdaten verwenden Fans von Sängern DieLochis Hashtags wie *#Lochinator* und Selbstbezeichnungen wie „größter Fan", um ihre Gruppenzugehörigkeit auszudrücken. Auch Creator*innen können spezifische Mittel wie Emojis oder Hashtags gezielt einsetzen, um diese als Erkennungssymbole ihrer Community zu etablieren, wie zum Beispiel das 🦄-Emoji, das im Fall der Creatorin BonnyTrash als Gruppenidentitätssymbol fungiert (siehe Beispiel 5). Auch die Kategorien *gratitude* und *sympathy* sind spezifischer ausgeprägt und treten nahezu ausschließlich in Kommentaren zu Videos auf, die entsprechende Emotionen hervorrufen. So finden sich Sympathiebekundungen in unseren Korpusdaten vor allem in Kommentaren zu einem Spendenaufruf für die Philippinen der YouTuberin Anna Maria Damm, während aufrichtiger Dank häufig in Kommentaren zu einem Fitness-Tutorial von Sophia Thiel geäußert wird. Die Kategorie *agreement* (1,2%) umfasst Flauschausdrücke, die als zustimmende Reaktionen auf flauschenthaltende Kommentare anderer Nutzer*innen fungieren und somit Teile unterstützender Diskurse markieren.

Die Verteilung der einzelnen Flauschtypen nach Creator*innen ist in Abbildung 2 (relative Häufigkeiten; siehe unten) sowie in Tabelle 2 (absolute Zahlen; siehe unten) dargestellt. Einige Flauschtypen kommen in allen Datensatzteilen vor und sind für jede kommentierende Community relevant. So sind *positive feedback*, *affection declaration* und *compliment* bei allen Creator*innen (wenn auch in unterschiedlichem Ausmaß) vorhanden; diese Flauschtypen können als communityübergreifender Flausch betrachtet werden. Darüber hinaus lassen sich communityspezifische Flausch-Tendenzen beobachten, wie z. B. ein besonders hoher Anteil an Komplimenten bei Nilam oder an *group membership* bei BonnyTrash. Zusätzlich gibt es videospezifischen Flausch, wie *gratitude* bei Sophia Thiel (als Dank für ein Fitness-Tutorial) oder *sympathy* bei Anna Maria Damm (ausgelöst durch den Spendenaufruf).

18 Da in dieser Darstellung nur die Gesamthäufigkeiten berücksichtigt werden, erfolgt keine Differenzierung der Werte nach den einzelnen Creator*innen. Diese allgemeine Übersicht soll übergeordnete Tendenzen aufzeigen.

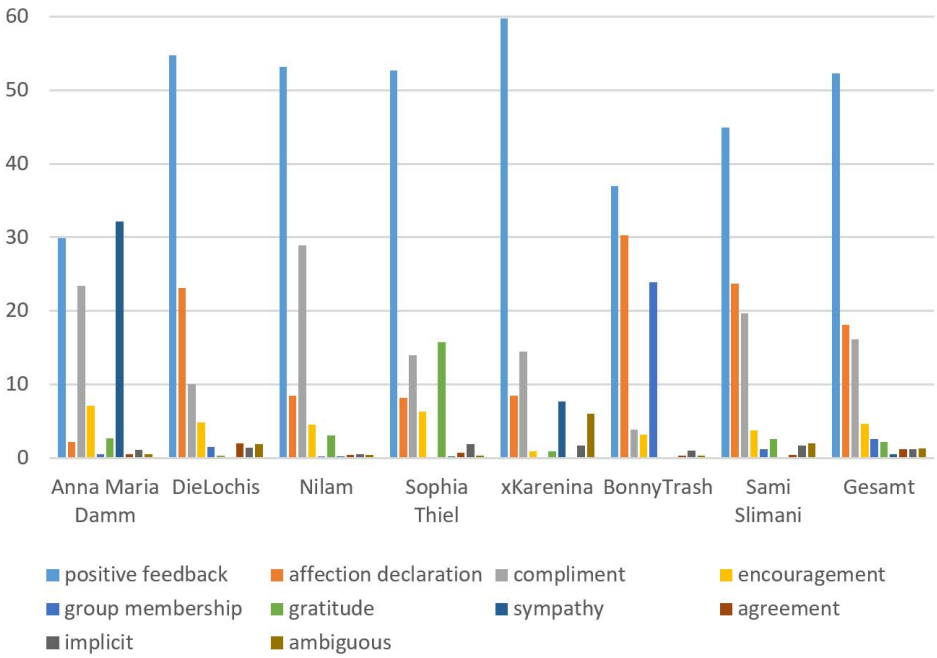


Abbildung 2: Relative Häufigkeitsverteilung der Flauschtypen nach Creator*innen (in %)

Besonders in den Datensatzteilen mit mehreren Videos, wie denen von DieLochis und Nilam, zeigen sich deutliche Unterschiede in der Verwendung von Flausch (z. B. im Hinblick auf das Vorkommen oder Ausbleiben von Danksagungen). Diese Unterschiede liefern Hinweise auf die jeweiligen Fan-Communities und ermöglichen eine tiefere Analyse, insbesondere in Bezug darauf, warum bestimmte Flauschtypen in den Kommentaren unterschiedlicher Kanäle häufiger vorkommen als andere. Darüber hinaus können die Unterschiede auf Besonderheiten einzelner Inhalte, wie beispielsweise beim Video von Anna Maria Damm, oder deren Qualität bzw. darauf, wie gut der Content bei den Nutzer*innen ankommt, zurückzuführen sein. Sie bieten daher Ansatzpunkte für weiterführende Analysen zu der Frage, welche Faktoren das Auftreten bestimmter Flauschtypen in Kommentaren beeinflussen. Wie die von uns verwendete Korpusdaten zeigen, können dabei auch linguistische Unterschiede sichtbar werden. Zum Beispiel kann Dankbarkeit bei DieLochis (15–16) und Nilam (17–18) in ihrer Funktion identisch sein, sich aber in der sprachlichen Ausprägung deutlich unterscheiden. Der Ausdruck von Dankbarkeit in der Community von DieLochis erscheint emotionaler und stärker involviert; so wird er in (15) indirekt über eine positive Zuschreibung realisiert und durch Emojis emotional verstärkt. In der Community von Nilam erfolgt er dagegen eher formelhaft und konventionalisiert, wie in (17), wo er explizit mittels eines formelhaften Ausdrucks realisiert wird (siehe auch Leuckert 2020 für linguistische Unterschiede zwischen Subreddit-Communities).

Creator*in	Flauschtyp										
	positive feedback	affection declaration	compliment	encouragement	group membership	gratitude	sympathy	agreement	implicit	ambiguous	Gesamt
Anna Maria Damm	55	4	43	13	1	5	59	1	2	1	184
DieLochis	6 131	2 592	1 124	534	170	34	15	225	161	218	11 204
Nilam	3 529	567	1 919	300	12	205	14	26	36	28	6 636
Sophia Thiel	690	107	183	82	1	206	3	9	25	4	1 310
xKarenina	70	10	17	1	0	1	9	0	2	7	117
Bonnytrash	564	461	58	49	364	2	0	5	15	5	1 523
Sami Slimani	364	192	160	30	10	21	1	3	14	16	811
Gesamt	11 403	3 933	3 504	1 009	558	474	101	269	255	279	21 785

Tabelle 2: Verteilung der Flauschtypen (absolute Häufigkeiten)

(15) Ihr seid die letzten die einem Hoffnung schenken 💞 🙏

(NDY/208)

(16) Ich danke euch für jedes noch so GEILES video

(NDY/186)

(17) vielen Dank für die heiteren Minuten daaruum !

(NDY/138)

(18) Danke für diese Info , ernsthaft :)

(NDY/250)

5 Klassifikationen positiver und negativer Sprache im Vergleich

Wie bereits erwähnt, stellt Flauschrede – ebenso wie andere Formen positiver Sprache – den Gegenpol zur Hassrede in der virtuellen Kommunikation dar. Während Hassrede darauf abzielt, Personen zu verletzen und deren Selbstbild zu destabilisieren, zielt positive Sprache darauf ab, das Selbstbild der Adressat*innen aufzubauen, zu stärken und zu unterstützen.

Im Folgenden beleuchten wir zunächst bestehende Klassifikationen positiver und negativer Sprache und setzen diese anschließend in Bezug zu unserer Klassifikation von Flausch.

5.1 Klassifikationen positiver Onlinesprache

Die bislang entwickelten Ansätze zur Analyse positiver Onlinesprache – insbesondere zur Hoffnungsrede und bestärkender Sprache (siehe Abschnitt 3.2) – beschränken sich auf binäre Klassifikationen von Social-Media-Beiträgen, wie YouTube-Kommentaren oder Reddit-Posts. Dabei wird erfasst, ob bestimmte Merkmale der jeweiligen Form positiver Sprache in einem Beitrag vorhanden sind, während feinere Differenzierungen, also welche Textteile tatsächlich positiv sind, unberücksichtigt bleiben.

So wird bestärkende Sprache in Njoo et al. (2023) danach klassifiziert, ob Reddit-Posts als bestärkend gelten oder nicht. Auch Ansätze zur Hoffnungsrede folgen überwiegend diesem binären Schema und kategorisieren vollständige Beiträge auf verschiedenen Social-Media-Plattformen hinsichtlich des Vorliegens entsprechender Merkmale (siehe u. a. Chakravarthi 2020; Malik et al. 2023; Nath/Singh/Gupta 2025).

Einige Studien nehmen differenziertere Klassifikationen vor und unterscheiden zwischen verallgemeinerter, realistischer und unrealistischer Hoffnung (Balouchzahi et al. 2023) oder zwischen Varianten von Hoffnungsrede, die Glaube, Wunsch oder Kontrafaktizität ausdrücken (da Silva et al. 2025). Ein kürzlich veröffentlichter Datensatz unterscheidet vier Arten von Hoffnungsrede auf Reddit und Twitter/X: Inspiration, Solidarität, Resilienz und Spiritualität (Zaghoauni/Biswas 2025). Ein weiterer Datensatz unterscheidet zwischen optimistischer, realistischer, inspirierender und schwindender Hoffnung sowie Hoffnungslosigkeit (Thenmozhi et al. 2026). Dennoch bleibt auch hier der gesamte Kommentar oder Post die zentrale, atomare Analyseeinheit. Darüber hinaus variieren die Datensätze in ihrer Konzeption und klassifizieren Untertypen der Hoffnungsrede auf unterschiedlichen inhaltlichen Ebenen, sodass keine einheitliche Vergleichbarkeit besteht.

Wie unsere YouTube-Daten zeigen, lassen sich ganze Kommentare nicht immer eindeutig als positiv oder negativ klassifizieren. Häufig tritt Flausch gemeinsam mit neutralen oder sogar abwertenden Formulierungen auf, wie nachfolgende Belege in (19) und (20) aus dem Jahr 2014 in unserem Datensatz illustrieren. Dabei beziehen sich positive und negative Elemente nicht zwangsläufig auf dieselbe Zielperson. Im Kommentar in (20) richten sich die abwertenden Formulierungen gegen andere Nutzer*innen, die den Content kritisieren, während zugleich Flausch gegenüber den Creator*innen enthalten ist. In (19) wird der Content der Creator*innen positiv bewertet, während zugleich auf anderen Content Bezug genommen und dieser negativ bewertet wird, was jedoch nicht mit Hassrede gleichzusetzen ist.

(19) cooles video !! einfach nur geil !!! die mit der Jenny istnicht gut !! hört euch mal beide an ... dann bemerktihr den unterschied

(20) haltet euro fresse die sind cool ihr dummen hater

(NDY/004)

Um dieser Komplexität gerecht zu werden, ist eine feingranulare Differenzierung notwendig, sowohl hinsichtlich verschiedener Flauschtypen als auch in Bezug auf die präzise sprachliche Abgrenzung relevanter Textsegmente innerhalb eines Beitrags. Daher erfolgt unsere Annotation nicht auf Ebene des gesamten Kommentars, sondern auf Ebene einzelner sprachlicher Segmente. Der hier vorgestellte Ansatz entwickelt somit eine Systematisierung positiver Sprache, die eine differenzierte Betrachtung ermöglicht. Diese kann außerdem als Ausgangspunkt für vergleichbare Analysen negativer Sprache dienen, zu denen bereits umfangreiche Arbeiten existieren, die jedoch nur teilweise eine ähnlich feine Granularität der Klassifikation erreichen (siehe Abschnitt 5.2). Auf dieser Grundlage ließe sich anschließend eine Gesamtbewertung eines Beitrags vornehmen, bei der verschiedene Formen positiver und negativer Sprache und gegebenenfalls auch neutrale Sprache berücksichtigt und deren Anteile in geeigneter Weise zusammengeführt bzw. verrechnet werden. Siehe hierzu beispielsweise die Arbeit von Sidarenka (2016), in der ein Sentiment-Annotationsschema entwickelt wird, das die Annotation sentimentrelevanter Textsegmente einschließlich ihrer Quelle und ihres Ziels (z. B. Person oder Objekt) berücksichtigt.

5.2 Klassifikationen von Hassrede

Wie bereits erwähnt, stellt Hassrede ein komplexes und vielschichtiges Phänomen dar (siehe u. a. Schmidt/Wiegand 2017; Fortuna/Nunes 2018; Poletto et al. 2021; Jaki 2023). Neben dem Begriff „Hassrede“ finden sich in der Fachliteratur weitere Bezeichnungen wie verletzende Sprache (*abusive language*; Fortuna et al. 2023), toxische Sprache (*toxic language* oder *toxicity*; Pavlopoulos et al. 2020), beleidigende Sprache bzw. Anstößigkeit (*offensive language* bzw. *offensiveness*; Aggarwal et al. 2019; Zampieri et al. 2019) sowie schädigende Sprache (*harmful speech*; Scheffler/Solopova/Popa-Wyatt 2021; 2024).

Poletto et al. (2021) zeigen, dass diese Konzepte zwar eng miteinander verwoben sind, sich jedoch nicht eins zu eins aufeinander abbilden lassen. Ein vergleichbares Bild ergibt sich in Abschnitt 3.2 im Hinblick auf Flauschrede und verwandte Phänomene: Auch im Bereich positiver expressiver Sprache existieren unterschiedliche Konzepte, die bestimmte Teilaspekte abbilden, sich jedoch in ihren Einzelheiten unterscheiden. Wie bei der positiven Sprache sind auch die meisten Klassifikationen von Hassrede und verwandten Phänomenen binär angelegt und bestimmen, ob eine Aussage Hassrede enthält oder nicht. Einige Ansätze erweitern dieses Schema um zusätzliche Kategorien, beispielsweise durch die Unterscheidung

zwischen obszönen (*obscene*) und beleidigenden (*offensive*) Hassausdrücken (Mubarak et al. 2017) oder zwischen Hassrede und Ausdrücken, die beleidigend sind, aber nicht als Hassrede gelten (*offensive but not hate speech*) (Martins et al. 2018). Darüber hinaus existieren Ansätze, die graduelle Abstufungen von Hassrede berücksichtigen, wie zum Beispiel sehr toxisch, toxisch, leicht toxisch und nicht toxisch (Kolhatkar et al. 2020) oder direkt aggressiv, indirekt aggressiv und nicht aggressiv (Kumar et al. 2018). Zudem werden häufig adressatenbezogene Differenzierungen vorgenommen, unter anderem entlang sozialer, religiöser, sexistischer oder rassistischer Dimensionen (Waseem/Hovy 2016) sowie zwischen aufhetzender, kommunaler und religiöser Hassrede (Ishmam/Sharmin 2019).

Wenige Klassifikationen ermöglichen eine feingliedrigere Typisierung. Ein Beispiel hierfür ist der GermEval 2019 Shared Task zur Identifikation beleidigender Sprache, der zwischen Beleidigung, Profanität und Missbrauch differenziert (Struß et al. 2019). Eine ähnliche Herangehensweise findet sich in einer fünfteiligen Taxonomie schädigender Sprache, die Unterkategorien umfasst (Scheffler/Solopova/Popa-Wyatt 2021). Diese Taxonomie ist nach ähnlichen funktionalen Aspekten aufgebaut wie die hier vorgeschlagene Kategorisierung von Flauschausdrücken. Sie umfasst eine Annotation verschiedener Formen negativer Sprache, darunter aufhetzende Sprache, abwertende Worte und Ausdrücke, beleidigende, missbräuchliche oder anstößige Äußerungen, ein- und ausgrenzende Sprache (In-/Out-Group) sowie Codes.

Unsere Klassifikation der Flauschausdrücke umfasst sieben Typen, die unterschiedliche sprachliche Strategien unterstützender Kommunikation abbilden. Ergänzend umfasst sie zusätzliche Klassen für impliziten Flausch (*implicit*), Zustimmung zu Flausch (*agreement*) sowie mehrdeutige Ausdrücke (*ambiguous*), die aufgrund fehlenden Kontexts nicht eindeutig als Flausch klassifiziert werden können. Während *implicit* und *agreement* je nach Forschungsinteresse weiter differenziert werden können, stellt sich für die Kategorie *ambiguous* die Frage nach einem angemessenen analytischen Umgang; sie kann entweder gesondert ausgewiesen oder von weiterführenden Analysen ausgeschlossen werden.

Flausch und Hass lassen sich – in all ihren Ausprägungen – als zwei Seiten expressiver Sprache verstehen; dabei weist Flausch eine mit Hass vergleichbare Komplexität auf. Um diese Komplexität zu erfassen – ähnlich wie in zahlreichen Studien zur negativen Sprache – verfolgen wir das Ziel, mit einer Definition und Klassifikation von Flauschrede einen Beitrag zu dieser Forschungslücke zu leisten. Die Entwicklung differenzierter Annotationsrichtlinien für sowohl virtuelle Flausch- als auch Hassrede bildet die Grundlage für eine um-

fassende empirische Untersuchung der Expressivität der Onlinesprache, die sowohl korpus- als auch computerlinguistische Methoden miteinbeziehen kann (vgl. Scheffler 2025). Während für Hassrede bereits zahlreiche Taxonomien existieren, wurde in diesem Beitrag eine systematische Klassifikation von Flauschausdrücken für die positive Seite expressiver Sprache vorgeschlagen.

6 Zusammenfassung und Ausblick

In virtuellen Räumen finden vielfältige Formen sprachlicher und nicht-sprachlicher Interaktion statt. Als Gegenpol zur viel untersuchten Hassrede führten wir den Begriff des virtuellen Flauschs ein und grenzten ihn von anderen Formen positiver Onlinekommunikation ab. Flausch ist ebenso komplex wie die expressive negative Sprache und umfasst sprachliche Ausdrücke sowie (nicht-sprachliche) Elemente wie Emojis und Hashtags, die der Unterstützung und Stärkung des positiven Selbstbildes der Adressat*innen dienen. Wir unterscheiden sieben Typen von Flausch, die klar voneinander abgrenzbar sind und eine systematische sowie differenzierte Analyse positiver Kommunikationspraktiken in virtuellen Gemeinschaften ermöglichen.

Im Rahmen einer empirischen Untersuchung analysierten wir das Vorkommen von Flauschausdrücken in Kommentaren zu deutschsprachigen YouTube-Videos aus dem NottDeuYTSch-Korpus. Etwa ein Drittel der Kommentare in unserem Datensatz enthält Äußerungen, die unserer Definition von Flausch entsprechen. Das legt nahe, dass Flausch in den YouTube-Kommentarbereichen der Videos unterschiedlicher Creator*innen und somit auch in diversen virtuellen Fan-Gemeinschaften einen relevanten Anteil der Kommunikation ausmacht, was zur Attraktivität und Beliebtheit des digitalen Austauschs beitragen kann.

Diese Studie verfolgte primär das Ziel, das Forschungsinteresse an der datenbasierten Analyse positiver expressiver Sprache zu fördern und neue methodische Perspektiven auf gesichtsunterstützende sprachliche Praktiken in virtuellen Räumen zu eröffnen. Wir fokussierten uns zunächst auf den Bereich der Jugendkultur auf YouTube. Eine Übertragung auf andere Domänen und Medien stellt daher einen wichtigen nächsten Schritt dar, da auch dort Flausch auftreten könnte und möglicherweise weitere Ausprägungen sowie Differenzierungen dieses Phänomens sichtbar werden. Darüber hinaus sehen wir Potenzial in einer kombinierten Betrachtung positiver und negativer Ausdrücke in Beiträgen. Dies könnte ein umfassenderes Bild sprachlicher Interaktionen auf unterschiedlichen Plattfor-

men und in verschiedenen Kontexten ermöglichen, indem die jeweiligen Anteile positiver und negativer Sprache berücksichtigt werden und dadurch Unterschiede in der sprachlichen Gestaltung unterschiedlicher Medien und Kommunikationsräume untersucht werden können.

Wie im Verlauf dieses Aufsatzes aufgeführt, bieten sich verschiedene Ansätze für weiterführende und vertiefende Studien an, wie beispielsweise Medium-vergleichende Analysen, vertiefte linguistische Untersuchungen oder die Analyse von Prozessen der Community- und Identitätsbildung. Des Weiteren hoffen wir, dass unsere Definition von Flauschrede sowie der dazugehörige Datensatz eine vielversprechende Grundlage für zukünftige (computer-)linguistische Untersuchungen positiver Interaktionen in virtuellen Lebenswelten bilden. Die segmentbasierte Annotation ermöglicht nicht nur quantitative Analysen des Auftretens unterschiedlicher Formen von Flausch, sondern kann auch als Ausgangspunkt für die Entwicklung und Evaluation automatischer Verfahren zur Erkennung und Differenzierung positiver sprachlicher Kommunikation dienen. Die Potenziale des Datensatzes für weiterführende computerlinguistische Anwendungen wurden bereits im Rahmen eines Shared Tasks demonstriert. Dabei wurden verschiedene Modelle entwickelt, die Flauschausdrücke automatisch erkennen können. Ferner eröffnet der annotierte Datensatz spezifischere Anwendungsmöglichkeiten, beispielsweise die Identifikation schmeichelnder Ausdrücke in den Ausgaben großer Sprachmodelle, um Phänomene wie Sycophancy systematisch zu erkennen und zu überwachen (vgl. Thelen et al. 2025). Im nächsten Schritt unserer Forschung erweitern wir den hier erstellten Datensatz und untersuchen die Identitäts- und Gruppenbildungsprozesse in verschiedenen virtuellen Fan-Communities. Dabei analysieren wir insbesondere, welche (linguistischen) Merkmale positiver Kommunikation für diese Prozesse relevant sind und wie sie sich zwischen Fan-Gemeinschaften verschiedener Creator*innen unterscheiden.

Förderung

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 506106235 – SFB 1567.

Literatur

Aggarwal, Piush/Horsmann, Tobias/Wojatzki, Michael/Zesch, Torsten (2019): LTL-UDE at SemEval-2019 Task 6: BERT and Two-Vote Classification for Categorizing Offensiveness. In: *Proceedings of the 13th International Workshop on Semantic Evaluation*. Association for Computational Linguistics, 678–682.
URL: https://aclanthology.org/S19-2121/

Ashraf, Noman/Zubiaga, Arkaitz/Gelbukh, Alexander (2021): Abusive language detection in YouTube comments leveraging replies as conversational context. In: *PeerJ Computer Science*, 7:e742.
DOI: https://doi.org/10.7717/peerj-cs.742

Balouchzahi, Fazlourrahman/Sidorov, Grigori/Gelbukh, Alexander (2023): PolyHope: Two-level hope speech detection from tweets. In: *Expert Systems with Applications*, 225 (120078), 1–13.
DOI: https://doi.org/10.1016/j.eswa.2023.120078

Beißwenger, Michael/Pappert, Steffen (2019): *Face work* mit Emojis. Was linguistische Analysen zum Verständnis sprachlichen Handelns in digitalen Lernumgebungen beitragen können. In: Beißwenger, Michael/Knopp, Matthias (Hg.): *Soziale Medien in Schule und Hochschule: Linguistische, sprach- und mediendidaktische Perspektiven*. Frankfurt: Peter Lang (Forum Angewandte Linguistik 63), 101–143.
URL: https://www.peterlang.com/view/title/68195

Bendel, Oliver (2019): *350 Keywords Digitalisierung*. Springer Gabler, Wiesbaden.
DOI: https://doi.org/10.1007/978-3-658-25823-8_3

Biri, Ylva (2023): Affective stance as positioning in an English-speaking online fan community. In: *Journal of Pragmatics* 215, 70–83.
DOI: https://doi.org/10.1016/j.pragma.2023.07.004

Brodnig, Ingrid (2016): *Hass im Netz: Was wir gegen Hetze, Mobbing und Lügen tun können*. Wien: Christian Brandstätter Verlag.

Bundesnetzagentur (2023): Nutzung von Online-Kommunikationsdiensten in Deutschland: Ergebnisse der Verbraucherbefragung 2023 (Kurzfassung).
URL: https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Digitales/OnlineKom/befragung_kurz23.pdf

Burnett, Heather (2017): Sociolinguistic interaction and identity construction: The view from game-theoretic pragmatics. In: *Journal of Sociolinguistics* 22 (1), 238–271.
DOI: https://doi.org/10.1111/josl.12229

Burnett, Heather (2019): Signalling games, sociolinguistic variation and the construction of style. In: *Linguistics and Philosophy* 42 (5), 419–450.

DOI: https://doi.org/10.1007/s10988-018-9254-y
Chakravarthi, Bharathi Raja (2020): HopeEDI: A Multilingual Hope Speech Detection Dataset for Equality, Diversity, and Inclusion. In: *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media*. Association for Computational Linguistics, 41–53.
URL: https://aclanthology.org/2020.peoples-1.5/
Chakravarthi, Bharathi Raja/Muralidaran, Vigneshwaran/Priyadharshini, Ruba/Cn, Subalalitha/McCrae, John/García, Miguel Ángel/Jiménez-Zafra, Salud María/Valencia-García, Rafael/Kumaresan, Prasanna/Ponnusamy, Rahul/García-Baena, Daniel/García-Díaz, José (2022): Overview of the Shared Task on Hope Speech Detection for Equality, Diversity, and Inclusion. In: *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics, 378–388.
DOI: https://doi.org/10.18653/v1/2022.ltedi-1.58
Chayko, Mary (2020): Rethinking Community in Communication and Information Studies: Digital Community and Community 'to Go'. In: Jansen, Bettina (Hg.): *Rethinking Community through Transdisciplinary Research*. Palgrave Macmillan, Cham, 99–110.
DOI: https://doi.org/10.1007/978-3-030-31073-8_6
Clausen, Yulia & Scheffler, Tatjana (2025): Annotating candy speech in German YouTube comments. In: *Proceedings of the 19th Linguistic Annotation Workshop (LAW-XIX-2025)*, Wien, Österreich. Association for Computational Linguistics, 264–269.
DOI: https://doi.org/10.18653/v1/2025.law-1.21
Clausen, Yulia/Scheffler, Tatjana/Wiegand, Michael (2025): Overview of the GermEval 2025 Shared Task on Candy Speech Detection. In: *Proceedings of the 21st Conference on Natural Language Processing (KONVENS 2025): Workshops*. HsH Applied Academics, 388–397.
URL: https://aclanthology.org/2025.konvens-2.32/
Cotgrove, Louis (2023): THE NOTTDEUYTSCH CORPUS: A corpus of German-language YouTube comments. In: *Korpora Deutsch als Fremdsprache* 3 (2), 225–229.
DOI: https://doi.org/10.48694/kordaf.3841
Cotgrove, Louis (2025): *Abogeil! The language of German teens on YouTube*. Leibniz-Institut für Deutsche Sprache, Mannheim.
DOI: https://doi.org/10.14618/amades-63
Darics, Erika (2010): Politeness in computer-mediated discourse of a virtual team. In: *Journal of Politeness Research*, 6(1), 129–150.
DOI: https://doi.org/10.1515/jplr.2010.007
Da Silva, Tulio Ferreira Leite/Aduna, Gonzalo Freijedo/Benamara, Farah/Mari, Alda/Li, Zongmin/Yue, Li/Su, Jian (2025): CDB: A

Unified Framework for Hope Speech Detection Through Counterfactual, Desire and Belief. In: *Findings of the Association for Computational Linguistics: NAACL 2025*. Association for Computational Linguistics, 4448–4463.
DOI: https://doi.org/10.18653/v1/2025.findings-naacl.252

García-Baena, Daniel/García-Cumbreras, Miguel Ángel/Jiménez-Zafra, Salud María/García-Díaz, José Antonio/Valencia-García, Rafael (2023): Hope speech detection in Spanish. In: *Lang Resources & Evaluation* 57, 1487–1514.
DOI: https://doi.org/10.1007/s10579-023-09638-3

Goffman, Erving (1967): *Interaction Ritual: Essays on Face-to-Face Behavior*. Anchor Books, Doubleday & Company, Inc, Garden City, New York.

Graham, Sage Lambert (2007): Disagreeing to agree: Conflict, (im)politeness and identity in a computer-mediated community. In: *Journal of Pragmatics 39* (4), 742–759.
DOI: https://doi.org/10.1016/j.pragma.2006.11.017

Ishmam, Alvi Md/Sharmin, Sadia (2019): Hateful Speech Detection in Public Facebook Pages for the Bengali Language. *18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, Boca Raton, FL, USA, 555–560.
DOI: https://doi.org/10.1109/ICMLA.2019.00104

Jaki, Sylvia (2023): Hate Speech in sozialen Medien: Ein Forschungsüberblick aus Sicht der Sprachwissenschaft. In: Jaki, Sylvia/Steiger, Stefan (Hg.): *Digitale Hate Speech: Interdisziplinäre Perspektiven auf Erkennung, Beschreibung und Regulation*, Berlin: J.B. Metzler.

Fortuna, Paula/Nunes, Sérgio (2018): A survey on automatic detection of hate speech in text. In: *ACM Computing Surveys (CSUR)* 51 (4), 1–30.
DOI: https://doi.org/10.1145/3232676

Fortuna, Paula/Soler-Company, Juan/Wanner, Leo (2023): Dataset annotation in abusive language detection. In: Strippel, Christian/Paasch-Colberg, Sünje/Emmer, Martin/Trebbe, Joachim (Hg.): *Challenges and perspectives of hate speech research*, 443–464.
DOI: https://doi.org/10.48541/dcr.v12.26

Hosseinmardi, Homa/Mattson, Sabrina Arredondo/Ibn Rafiq, Rahat/Han, Richard/Lv, Qin/Mishra, Shivakant (2015): Analyzing Labeled Cyberbullying Incidents on the Instagram Social Network. In: Liu, Tie-Yan/Scollon, Christie Napa/Zhu, Wenwu (Hg.): *Social Informatics. SocInfo 2015. Lecture Notes in Computer Science* 9471, 49–66. Springer, Cham.
URL: https://link.springer.com/chapter/10.1007/978-3-319-27433-1_4

Kapil, Prashant/Ekbal, Asif (2024): A Survey on Combating Hate Speech through Detection and Prevention in English. In: *Proceedings of the 21st International Conference on Natural Language Processing (ICON)*. NLP Association of India (NLPAI), 485–501.
URL: https://aclanthology.org/2024.icon-1.57/

Kersten, Saskia/Lotze, Netaya (2019): Creating a Self-Image: Face-Work and Identity Construction Online. In: *Journal für Medienlinguistik* 2 (2), 123–156.
DOI: https://doi.org/10.21248/jfml.2019.20

Khalid, Osama/Srinivasan, Padmini (2020): Style Matters! Investigating Linguistic Style in Online Communities. In: *Proceedings of the International AAAI Conference on Web and Social Media* 14 (1), 360–369.
DOI: https://doi.org/10.1609/icwsm.v14i1.7306

Klie, Jan-Christoph/Bugert, Michael/Boullosa, Beto/Eckart de Castilho, Richard/Gurevych, Iryna (2018): The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. In: *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*. Association for Computational Linguistics, 5–9.
URL: https://aclanthology.org/C18-2002/

Klüwer, Tina (2011): "I Like Your Shirt" – Dialogue Acts for Enabling Social Talk in Conversational Agents. In: Vilhjálmsson, Hannes Högni/Kopp, Stefan/Marsella, Stacy/Thórisson, Kristinn R. (Hg.): *Intelligent Virtual Agents. IVA 2011. Lecture Notes in Computer Science* 6895. Springer, Berlin, Heidelberg, 14–27.
DOI: https://doi.org/10.1007/978-3-642-23974-8_2

Kolhatkar, Varada/Wu, Hanhan/Cavasso, Luca/Shukla, Kavan/Taboada, Maite (2020): The SFU Opinion and Comments Corpus: A Corpus for the Analysis of Online News Comments. In: *Corpus Pragmatics* 4, 155–190.
DOI: https://doi.org/10.1007/s41701-019-00065-w

Kumar, Ritesh/Ojha, Atul Kr./Malmasi, Shervin/Zampieri, Marcos (2018): Benchmarking Aggression Identification in Social Media. In: *Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018)*. Association for Computational Linguistics, 1–11.
URL: https://aclanthology.org/W18-4401/

Leuckert, Sven (2020): Rethinking Community in Linguistics: Language and Community in the Digital Age. In: Jansen, Bettina (Hg.): *Rethinking Community through Transdisciplinary Research*. Palgrave Macmillan, Cham, 111–125.
DOI: https://doi.org/10.1007/978-3-030-31073-8_7

Maíz-Arévalo, Carmen/García-Gómez, Antonio (2013): 'You look terrific!' Social evaluation and relationships in online compliments. In: *Discourse Studies* 15 (6), 735–760.
DOI: https://doi.org/10.1177/1461445613490011

Malik, Muhammad Shahid Iqbal/Nazarova, Anna/Jamjoom, Mona Mamdouh/Ignatov, Dmitry I. (2023): Multilingual hope speech detection: A Robust framework using transfer learning of fine-tuning RoBERTa model. In: *Info & Claims* 35 (8).
DOI: https://doi.org/10.1016/j.jksuci.2023.101736

Martins, Ricardo/Gomes, Marco/Almeida, Jose Joao/Novais, Paulo/Henriques, Pedro (2018): Hate Speech Classification in Social Media Using Emotional Analysis, *7th Brazilian Conference on Intelligent Systems (BRACIS)*, 61–66.
DOI: https://doi.org/10.1109/BRACIS.2018.00019

Merten, Marie-Luis (2022): Ritualisierte Anschlusskommunikation auf Instagram: Formelhafte Komplimente und ihre Erwiderung als verdichtete Online-Interaktion. In: *Deutsche Sprache* 4 (4), 309–334.
DOI: https://doi.org/10.37307/j.1868-775X.2022.04.03

Mubarak, Hamdy/Darwish, Kareem/Magdy, Walid (2017): Abusive Language Detection on Arabic Social Media. In: *Proceedings of the First Workshop on Abusive Language Online*, Association for Computational Linguistics, 52–56.
DOI: https://doi.org/10.18653/v1/W17-3008

Nath, Tanusree/Singh, Vivek Kumar/Gupta, Vedika (2025): Bong-Hope: An annotated corpus for Bengali hope speech detection. In: *International Journal of Information Technology* 17, 2523–2531.
DOI: https://doi.org/10.1007/s41870-025-02484-2

Njoo, Lucille/Park, Chan/Stappart, Octavia/Thielk, Marvin/Chu, Yi/Tsvetkov, Yulia (2023): TalkUp: Paving the Way for Understanding Empowering Language. In: *Findings of the Association for Computational Linguistics: EMNLP 2023*. Association for Computational Linguistics, 9334–9354.
DOI: https://doi.org/10.18653/v1/2023.findings-emnlp.625

Palakodety, Shriphani/KhudaBukhsh, Ashiqur R./Carbonell, Jaime G. (2020): Hope Speech Detection: A Computational Analysis of the Voice of Peace. *24th European Conference on Artificial Intelligence – ECAI 2020*.
URL: https://ecai2020.eu/papers/869_paper.pdf

Pavlopoulos, John/Sorensen, Jeffrey/Dixon, Lucas/Thain, Nithum/Androutsopoulos, Ion. (2020): Toxicity Detection: Does Context Really Matter? In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 4296–4305.

DOI: https://doi.org/10.18653/v1/2020.acl-main.396
Poletto, Fabio/Basile Valerio/Sanguinetti, Manuela/Bosco, Cristina/Patti, Viviana (2021): Resources and benchmark corpora for hate speech detection: a systematic review. In: *Lang Resources & Evaluation* 55, 477–523.
DOI: https://doi.org/10.1007/s10579-020-09502-8
Placencia, María Elena/Lower, Amanda (2017): Compliments and compliment responses. In: Hoffmann, Christian/Bublitz, Wolfram (Hg.): *Pragmatics of Social Media*. Berlin, Boston: De Gruyter Mouton, 633–660.
DOI: https://doi.org/10.1515/9783110431070-023
Placencia, María Elena/Eslami, Zohreh R. (2020): *Complimenting Behavior and (Self-)Praise across Social Media: New contexts and new insights*. John Benjamins Publishing Company.
DOI: https://doi.org/10.1075/pbns.313
Rheingold, Howard (1994): *Virtuelle Gemeinschaft. Soziale Beziehungen im Zeitalter des Computers*. Bonn u. a.: ADDISON-WESLEY.
Rieger, Stefan/Schäfer, Armin/Tuschling, Anna (2021): *Virtuelle Lebenswelten: Körper – Räume – Affekte*, Berlin, Boston: De Gruyter.
DOI: https://doi.org/10.1515/9783110638127
Scheffler, Tatjana (2025): Expressivity and Computational Linguistics. In: Gutzmann, Daniel/Turgay, Katharina (Hg.): *The Oxford Handbook of Expressivity*. Oxford: Oxford University Press.
Scheffler, Tatjana/Solopova, Veronika/Popa-Wyatt, Michaela (2021): The Telegram Chronicles of Online Harm. In: *Journal of Open Humanities Data* 7 (8).
DOI: https://doi.org/10.5334/johd.31
Scheffler, Tatjana/Solopova, Veronika/Popa-Wyatt, Michaela (2024): Verbreitungsmechanismen schädigender Sprache im Netz: Anatomie zweier Shitstorms. In: Gaderer, Rupert/Grömmke, Vanessa (Hg.): *Hass teilen: Tribunale und Affekte virtueller Streitwelten*. Bielefeld: transcript Verlag, 199–224.
DOI: https://doi.org/10.1515/9783839471951-010
Schmidt, Anna/Wiegand, Michael (2017): A Survey on Hate Speech Detection using Natural Language Processing. In: *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*. Association for Computational Linguistics, 1–10.
DOI: https://doi.org/10.18653/v1/W17-1101
Sidarenka, Uladzimir (2016): PotTS: The Potsdam Twitter Sentiment Corpus. In: Nicoletta Calzolari et al. (Hg.): *Proceedings of the Ninth International Conference on Language Resources and*

Evaluation (LREC 2016). European Language Resources Association (ELRA), 1133–1141.
URL: https://aclanthology.org/L16-1181/

Struß, Julia Maria/Siegel, Melanie/Ruppenhofer Josef/Wiegand, Michael/Klenner, Manfred (2019): Overview of GermEval Task 2, 2019 Shared Task on the Identification of Offensive Language. In: *Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019)*, 354–365.
URL: https://corpora.linguistik.uni-erlangen.de/data/konvens/proceedings/papers/germeval/GermEvalSharedTask2019Iggsa.pdf

Suler, John (2004): The Online Disinhibition Effect. In: *CyberPsychology & Behavior* 7 (3), 321–326.
DOI: https://doi.org/10.1089/1094931041291295

Thelen, Christian Rene/Blaneck, Patrick Gustav/Bornheim, Tobias/Grieger, Niklas/ Bialonski, Stephan (2025): AIxcellent Vibes at GermEval 2025 Shared Task on Candy Speech Detection: Improving Model Performance by Span-Level Training. In: *Proceedings of the 21st Conference on Natural Language Processing (KONVENS 2025): Workshops*. Hannover, Germany. HsH Applied Academics, 398–403.
URL: https://aclanthology.org/2025.konvens-2.33/

Thenmozhi, Durairaj/Shetty, Rathnakar P./Hegde, Parameshwar R./M D, Anusha/Adyanthaya, Raksha/Aljunid, Mohammed Fadhel/Kumaresan, Prasanna Kumar/Chakravarthi, Bharathi Raja. (2026): Findings of the Shared Task on Hope Speech Detection in Tulu. In: *Proceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Waseem, Zeerak/Hovy, Dirk (2016): Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter. In: *Proceedings of the NAACL Student Research Workshop.* Association for Computational Linguistics, 88–93.
DOI: https://doi.org/10.18653/v1/N16-2013

Wiegand, Michael/Siegel, Melanie/Ruppenhofer, Josef (2019): Overview of the GermEval 2018 Shared Task on the Identification of Offensive Language. In: *Proceedings of GermEval 2018, 14th Conference on Natural Language Processing (KONVENS 2018)*, 1–10.
URL: https://www.lsv.uni-saarland.de/wp-content/publications/2018/germeval2018_wiegand.pdf

Zaghoauni, Wajdi/Biswas, Md. Rafiul (2025): EmoHopeSpeech: An Annotated Dataset of Emotions and Hope Speech in English and Arabic. In: *Proceedings of the 15th International Conference on*

Recent Advances in Natural Language Processing - Natural Language Processing in the Generative AI Era, Association for Computational Linguistics, 1406–1412.
URL: https://aclanthology.org/2025.ranlp-1.162/

Zampieri, Marcos/Malmasi, Shervin/Nakov, Preslav/Rosenthal, Sara/Farra, Noura/Kumar, Ritesh (2019): Predicting the Type and Target of Offensive Posts in Social Media. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Band 1 (Long and Short Papers). Association for Computational Linguistics, 1415–1420.
URL: https://aclanthology.org/N19-1144/