

## 2 Hate Speech

„Wir TRAUERN NICHT sondern wir FEIERN ES! Nur ein Flüchtling, ein Flüchtling ist zu wenig: Das Meer hat schon mehr Flüchtlinge geschluckt!“ (in Hasselmann 2015), „Türken sollten die zwei Eulen ficken bis sie normal werden.“<sup>1</sup> In den sozialen Netzwerken wettern private und öffentliche Personen zunehmend ungehemmt gegen Flüchtlinge, MuslimInnen, Homosexuelle, AusländerInnen oder Frauen. Prominente Akteure befeuern ihrerseits mit unsäglichen Statements in der Öffentlichkeit diese feindliche Stimmungsmache. Sogenannte Hasspostings geraten dabei verstärkt in den Fokus öffentlicher Aufmerksamkeit und deren VerfasserInnen finden sich häufig vor Gericht wieder.

Auch wenn die Auseinandersetzung mit Hasspostings zur Folge hatte, die Debatte um *Hate Speech* neu anzufachen und einem neuen, breiteren Lernprozess im Umgang damit anzustoßen, bringt dies Entwicklung auch negative Folgen mit sich: Der zuvor schon polysemische Begriff verliert noch weiter an Konturen.

Der Begriff Hassposting steht im Alltagsverständnis für eine Reihe von Formen symbolischer Gewalt, wie *Bullying*, *Stalking*, Aufruf zu Straftaten, Verleumdung u.a. (Gagliardone et al. 2015, S. 7). „Alte Bekannte“ wie Beschimpfung, Drohung oder Aufrufe zur Gewalttaten, egal ob gruppenbezogen oder nicht, bekommen jetzt das Label *Hate Speech*.

Eine solche Unschärfe ist nicht folgenlos. Wie Max Weber schon warnte, „(...) der Gebrauch der undifferenzierten Kollektivbegriffe, mit denen die Sprache des Alltags arbeitet, ist stets Deckmantel von Unklarheiten des Denkens oder Wollens“. Es handelt sich aber um mehr als eine unabsichtliche Ungenauigkeit. Diese Undifferenziertheit kann auch als Ziel dienen:

---

<sup>1</sup> Kommentar eines Nutzers, veröffentlicht auf der Facebook-Seite „Unsere Blaue Seite“ am 5. April 2016.

Sie ist auch „(...) oft genug das Werkzeug bedenklicher Erschleichungen, immer aber ein Mittel, die Entwicklung der richtigen Problemstellung zu heben“ (Weber 1988, S. 212). Ein konkretes Beispiel dafür ist die Anwendung des Ausdruckes während des Wahlkampfes in Kenia 2013, als PolitikerInnen sich gegenseitig beschuldigten, *Hate Speech* zu betreiben (Benech 2014a).

Der Mangel eines einheitlichen Begriffes erschwert zudem eine juristische Regulierung des Phänomens (Matsuda 1989), hat aber auch Folgen für die Wissenschaft, denn eine klare Definition ist sowohl für die Theoriebildung als auch für die empirische Erfassung des Phänomens unerlässlich (Saleem et al. 2016).

Ziel dieses Kapitels ist es, *Hate Speech* zu definieren. Zuerst wird analysiert, warum *Hate Speech* überhaupt wichtig ist. Danach wird begründet, warum juristische Definitionen nicht konstitutiv für *Hate Speech* sind.

Im Anschluss daran wird eine Definition von *Hate Speech* entwickelt und geklärt, woraus *Hate Speech* besteht und welche Wesensmerkmale es aufweist. Danach werden die unterschiedlichen Spielarten von *Hate Speech* behandelt. Zuletzt wird diese diskriminierende Kommunikationsform anhand der Wesensmerkmale von anderen Formen symbolischer Abwertung wie Beschimpfung, Beleidigung, Fauxpas, *Political Incivility* und Rechtspopulismus abgegrenzt. Zuletzt werden mögliche Antworten auf *Hate Speech* diskutiert (*Counter Speech*).

## **2.1   Warum ist Hate Speech wichtig?**

*Hate Speech* wird oft als möglicher Auslöser offener Gewalt wie *Hate Crimes* oder einem Genozid angesehen. Mal wird die Relevanz betont, in dem einen Zusammenhang angenommen wird, mal wird diese angezweifelt, da kein direkter Zusammenhang nachgewiesen werden konnte (Ferguson 2016; Schoemaker und Stremlau 2014)

Beide konträren Positionen gehen von den gleichen Grundannahmen aus:

a) die Wirkung von *Hate Speech* wird allein und ausschließlich auf offene Gewalt reduziert und b) *Hate Speech* wird allein und ausschließlich als Ursache bzw. Auslöser betrachtet. Was aber sagt die Forschung dazu?

Der erste Punkt bei dieser Debatte in der Wissenschaft liegt in der Notwendigkeit zu definieren, was „Schaden“ (*Harm*) eigentlich bedeutet:

„(...) even among those who agree that speech harms, there is much debate about the precise relationship about the speech and the harm. Again, getting clear on this question is important for the same reasons that identifying the harms is important. First, the kind of empirical evidence that is needed to establish that speech of a given kind is harmful will vary greatly according to how the speech brings the harm. Evidence that speech of that kind constitutes harm by, for example, enacting a harmful policy will be quite different from evidence that the speech causes harm by, for example, altering the hearer's beliefs via persuasion. Second, what ought to be done about the harms depends greatly on how they are created.“ (Maitra und McGowan 2013, S. 8)

Die Folgen von *Hate Speech* müssen sich also nicht auf rechtsextreme offene Gewalttaten und Genozide beschränken. In diesem Zusammenhang unterscheiden Maitra und McGowan (2013) *Constitutive* und *Consequential Harms*. Ersteres wird durch *Hate Speech* selbst verursacht, während letzteres sich aus *Hate Speech* ergibt.

Der Schaden kann zudem individuell oder sozial sein. Zu den persönlichen Folgen von *Hate Speech* gehören psychologische Leiden, wie z.B. Depressionen oder Angstzustände. Soziale Schäden sind z.B. Einschränkungen der Bewegungs- und Versammlungsfreiheit (Matsuda 1989). *Hate Speech* kann auch gesellschaftliche Folgen wie Desintegration zeitigen und politische Konsequenzen für demokratische Deliberationsprozesse haben.

*Consequential Harms* können folgende Formen annehmen (Maitra und McGowan 2013):

- RezipientInnen negativer Stereotype zu überzeugen, was zu schädlichen Verhalten führen kann
- die Wahrnehmung der RezipientInnen von einer Gruppe negative zu prägen
- Diskriminierung zu normalisieren
- Rezipienten dazu zu bringen, diskriminierendes Verhalten nachzuahmen

Abgesehen von solchen symbolischen, nicht-physischen Schäden, kann *Hate Speech* durchaus auch im Zusammenhang mit offener Gewalt wie im Fall ethnischer Konflikten oder Genoziden vorkommen (Frère 2010; Benesch 2011; Gichuhi Kimotho und Nyaga 2016; Yanagizawa-Drott 2014; Tse-sis 2002; Gustafsson 2016; Straus 2007).

Einmal zusammengetragen, welche Schäden *Hate Speech* potentiell verursachen kann, stellt sich die Frage nach der empirischen Evidenz dar. Diesbezüglich existiert bereits eine große Vielfalt von Studien zu Diskriminierung allgemein, zu Medienwirkungen und spezifisch zu *Hate Speech*. Die nachfolgende Darstellung erhebt keinen Anspruch auf Vollständigkeit. Das Ziel dieses Exkurses beschränkt sich darauf, anhand eines Auszuges der empirischen Studien zum Thema zu verdeutlichen, warum *Hate Speech* als Forschungsgegenstand relevant ist.

Bezüglich der persönlichen, individuellen Folgen zeigen wissenschaftliche Studien empirische Evidenzen hinsichtlich der Wirkung von Diskriminierung im Allgemeinen sowie von symbolischer Diskriminierung und *Hate Speech* im Besonderen.

Hinsichtlich der Wirkung von Diskriminierung insgesamt stellten Bennett et al. (2005) mittels einer multivariaten Analyse fest, dass junge AfroamerikanerInnen, die über rassistische Erfahrungen berichten, mehr rauchen als andere ProbandInnen (vgl. auch Anderson 2013). Das gleiche erfolgt mit Maoris in Neuseeland (Harris et al. 2006). Meyer identifizierte anhand

der empirischen Evidenz zur Folge der Diskriminierung von Homosexuellen und Transsexuellen sogar einen sogenannten *Minority Stress*, der durch unterschiedliche Formen der Diskriminierung zu seelischen Problemen führt, inklusive einer Erhöhung der Selbstmordrate (Meyer 2003).

Hinsichtlich der individuellen Wirkung von *Hate Speech* stellten Eisenberg et al. (2003, S. 353) fest, dass verbale Belästigungen wie *Bullying* und *Hate Speech* zwischen Teenagerinnen aufgrund eines tatsächlichem oder zugeschriebenem Übergewichts mit geringerem Selbstbewusstsein, Depressionen und Suizidversuchen korrelierten.

Allerdings sind die individuellen Schäden von *Hate Speech* für diejenigen, die nicht als Zielscheibe dienen, oft nicht sichtbar. Das liegt sowohl an der Reaktion der angefeindeten Individuen als auch an der Nicht-Betroffenheit der anderen Gruppen.

Leets (2002, S. 353) konfrontierte eine Gruppe jüdischer StudentInnen und eine Gruppe homosexueller StudentInnen mit *Hate Speech*-Botschaften und erfasste empirisch ihre Reaktion nach den sogenannten *Harmful Speech Model*. Die Forscherin stellte fest, dass die kurzfristige Reaktion der Befragten bei 53 Prozent der jüdischen und bei 86 Prozent der homosexuellen StudentInnen emotional war, die meisten aber nicht darauf antworten würden: „The majority of participants (83%) viewed silence (...) as taking the higher moral ground“ (Leets 2002, S. 356). Auch wenn dieses eine nachvollziehbare Reaktion sein mag, könnte es Folgen für die Wahrnehmung der Schäden von *Hate Speech* für diejenige haben, die nicht als Zielscheibe fungieren (vgl. dazu McClelland und Hunter 1992).

Cowan und Hodge (1996) stellten ProbandInnen unterschiedliche Szenarien von *Hate Speech* vor und beobachteten deren Einschätzung der Anstößigkeit der Botschaften. Weiße Frauen empfanden die Inhalte als aggressiver als weiße Männer. Kein Unterschied wurde diesbezüglich zwischen Nicht-weißen Männern und Nicht-weißen Frauen beobachtet. Außerdem wurde ethnisches *Hate Speech* als verletzender eingeschätzt, als

dasjenige gegen Frauen oder Homosexuelle. Die Schlussfolgerung der Autorinnen lautet: „(...) speech is in the ear of the beholder, especially in terms of its offensiveness“ (Cowan und Hodge 1996, S. 367). Diejenigen, die nicht betroffen sind, empfinden solche Botschaften ebenfalls nicht als verletzend (McClelland und Hunter 1992).

Die Folgen von *Hate Speech*, insbesondere im Fall von Face-to-Face-*Hateful Speech* wurden bereits intensiv empirisch untersucht. Solche Studien bieten eine breite Palette empirischer Befunde. Bei medialen *Hate Speech*, so wie bei der Wirkung der Medien in Konflikten, ist das Bild komplexer und weniger eindeutig (Schoemaker und Stremlau 2014; Ferguson 2016; Leader Maynard und Benesch 2016) .

Was Diskriminierung und Medienwirkung betrifft, weist die Studie von Gilliam und Iyengar (2000) darauf hin, dass mediale Repräsentationen zu Inferenzprozesse bei den ZuschauerInnen führen. In einem Experiment simulierten die Forscher einen Beitrag in der lokalen Fernsehberichterstattung. Die US-amerikanische lokale Fernsehberichterstattung ist nicht nur eine der wichtigsten Informationsquellen der Bevölkerung, sondern auch stark durch Berichte über Kriminalität geprägt, die mehr als 75 Prozent der Berichte ausmachen können. In mehr als der Hälfte davon wird eine explizite Referenz bezüglich der ethnischen Zugehörigkeit der TäterInnen gegeben.

Gilliam und Iyengar (2000) bereiteten unterschiedliche Versionen des gleichen Berichts über Kriminalität für ihre ProbandInnen vor. In einer davon war der Täter schwarz, in der anderen ein Weißer und in der letzten wurde der Täter nicht gezeigt. 60 Prozent von denen, die die Version ohne Hinweis auf den Täter gesehen haben, behaupteten ihn gesehen zu haben, 70 Prozent davon identifizierten ihn zudem als schwarz (Gilliam und Iyengar 2000). Im Fall von Islamophobie stellten Eyssel, Geschke und Frindte (2015) einen Zusammenhang zwischen Islamophobie und Konsum von Privatfernsehen fest.

So wie bei der Studie von Cowan und Hodge (1996) zeigten Iyengar und Kinder (2010) in eine weitere Studie, dass die Wahrnehmung von Medieninhalten sich je nach Gruppe stark unterscheidet. Schwarze Menschen halten die Streitfrage um rassistische Diskriminierung für wichtiger als weiße (Iyengar und Kinder 2010, S. 32). Das deutet nicht nur auf das Potential von *Hate Speech* hin, Gesellschaften zu desintegrieren, sondern auch, warum dieser Prozess (für die „Mehrheit“) unbemerkt voranschreitet.

Bezüglich der sozialen Folgen von medialem *Hate Speech* befragten Frindte, Schurz und Roth (2013) deutsche und Nicht-deutsche MuslimInnen vor und nach der Publikation des Buches *Deutschland schafft sich ab* von Thilo Sarrazin, welche eine Deutschlandweite Polemik über muslimische MigrantInnen auslöste, nach deren Einstellung zueinander. Die Anzahl der Interviewten vor und nach der Veröffentlichung der Studien betrugen 700 und 350. Da es sich nicht um identische Personen, sondern um unabhängige Stichproben handelte, kann keine eindeutige kausale Wirkung der Buchveröffentlichung nachgewiesen werden. Allerdings lassen die Ergebnisse vermuten, dass die Integrationsbereitschaft der MuslimInnen mit der Kontroverse insgesamt gesunken ist. So gaben die Befragten vor der Sarrazin-Polemik eher zu, private Kontakte mit Deutschen zu pflegen als danach. Auch Vorurteile gegen den „Westen“ sind nach der Polemik gestiegen (Frindte et al. 2013).

Was die politischen Folgen von *Hate Speech* betrifft, liefert die Forschung um die Online-Deliberation einige Hinweise. So stellte Rußman fest, dass in kleinen *Facebook-Communities* eine gesteigerte Bereitschaft zu beobachten ist, selbst *Posts* und/oder Kommentare zu verfassen und zwar wenn sich die politischen Akteure in ihrer Online-Kommunikation an die Prinzipien einer verständigungsorientierten Kommunikation halten (Russmann 2015, S. 193).

Jenseits der symbolischen Folgen wird der Zusammenhang zwischen *Hate Speech* und offener Gewalt wie im Fall eines Genozids intensiv erforscht.

Derartige Untersuchungen beschäftigen sich mit einer spezifischen Form von *Hate Speech* – dem sogenannten *Dangerous Speech*.

Der bekannteste Fall ist wahrscheinlich der Genozid in Ruanda 1994. Das Radio spielt hier sowohl bei der Anstiftung von Hass zwischen den Gruppen als auch bei der Koordinierung der kollektiven Gewalt bzw. Ermordung der Minderheit eine entscheidende Rolle (Frère 2010). Das Radio war das Leitmedium des Landes gewesen. Eine Presse existierte zwar, spielte aber aufgrund der hohen Rate an Analphabeten kaum eine Rolle (Yanagizawa-Drott 2014; Straus 2007).

Das halbprivate Radio *Television Libre des Mille Collines* (RTLM) stiftete zum Hass an, indem es die Tutsi-Minderheit als Kakerlaken bezeichnete und klar machte, dass die Regierung keine Absicht habe, diese Minderheit vor Angriffen zu beschützen (Yanagizawa-Drott 2014). Auch bei der Koordinierung des Massenmordes war es aktiv: So wurden z.B. nicht nur Namen, sondern auch die Autokennzeichen der Mordbedrohten Personen im Radio vorgelesen (Straus 2007, S. 612), damit diese nicht nur persönlich identifiziert werden, sondern auch nicht mit dem Auto entkommen konnten (Benesch 2004, S. 62). Der Fall sorgte sogar dafür, dass JournalistInnenen durch das *United Nation International Criminal Tribunal for Rwanda* (CTR) wegen Anstiftung zum Genozid verurteilt wurden (Straus 2007).

Yanagizawa-Drott (2014) konnte zeigen, dass es einen Zusammenhang zwischen der Reichweite des Empfanges des „*Hate Radios*“ RTLM und der Anzahl, der im Rahmen des Genozides angeklagten MörderInnen gab. Das galt insbesondere im Fall kollektiver Gewalt (etwa durch Milizen). Der Autor kam daher zur Schlussfolgerung, dass die Rolle der Medien, vor allem bei der Koordination der kollektiven Gewalt, entscheidend war:



„The counterfactual estimates suggest that approximately 10% of overall participation can be attributed to the radio station's broadcasts, and almost one-third of the violence by militias and other armed groups.“ (Yanagizawa-Drott 2014, S. 1989)

In der multivariate Analyse stellte der Forscher zudem fest, dass die Medienwirkung nicht nur bei individueller Gewalt, sondern auch bei höherem Bildungsniveau sank.

Die Studie von Yanagizawa-Drott (2014) war schon vielfach Gegenstand der Kritik, insbesondere wegen des aufgestellten linearen Zusammenhangs zwischen Medien und Gewalt. Straus (2007) untersuchte ebenfalls den Zusammenhang zwischen den Morden während des Genozids in Ruanda und dem Radiokonsum anhand einer Reihe von Faktoren wie Reichweite der Übertragung, Zeit, Inhalte der Botschaften und Interviews mit HörerInnen und TäterInnen.

Er stellte einen existierenden, aber differenzierten Einfluss von *Hate Speech* auf die Gewalttaten fest. Ein Ergebnis war, dass die Übertragung nicht mit dem Beginn des Genozides korrelierte. Es gibt jedoch eine Korrelation zwischen den Gewalttaten und den Aufrufen dazu im Radio, die sogar Namen und Adressen der Opfer enthielten. Das gilt für die Hauptstadt und deren Umgebung.

Was der Inhalte der Botschaften betrifft, zeigt sich, dass offene Aufrufe zu Gewalttaten im Radio häufiger übertragen wurden, als die Anzahl der Morde sank. Allgemein verschärfen sich die Hassbotschaften nachdem die Hochphase des Völkermords vorbei war.

Die Ergebnisse von Strauss (2007) zeigen zudem wen solche Botschaften mobilisierten. Anhand der Interviews mit TäterInnen und HörerInnen ließ sich feststellen, dass *Hate Speech* hauptsächlich „Hardliner“ bzw. Schlüsselpersonen zu Gewalt mobilisierte:

„The perpetrators who say radio incited them were more likely to commit more violence and to be leaders of the killing than those who said radio did not incite them.“ (Straus 2007, S. 628).

Für die TäterInnen, die keine Führungspositionen bei der Gewaltaktionen innehatten, war eher der direkte Kontakt, normalerweise mit jemanden aus der Elite, einer Gruppe oder einem gewalttätigen Person das entscheidende Kriterium für die Teilnahme am Genozid.

Zusammenfassend lässt sich sagen, dass zahlreiche Studien eine Wirkung von *Hate Speech* nachweisen. Diese unterscheidet sich aber wesentlich zwischen den Studien und hängt nicht nur vom Kontext, sondern auch von der Art von *Hate Speech* ab. Auch was mediale Formen betrifft, kann eine Wirkung beobachtet werden. Allerdings ist diese nur eine unter mehreren Einflussfaktoren. Ob Medien der determinierende Faktor sind, lässt sich pauschal nicht beantworten.

Bei *Hate Speech*-Debatten wird aber häufig übersehen, dass diese nicht nur als Ursache, sondern auch als Indikator fungieren kann. So können z.B. *Derogatory Group Labels* auf soziale Ausgrenzung hinweisen. Mullen und Rice (2003) stellten einen Zusammenhang zwischen ethnischer Segregation am Wohnort und Arbeitsmarkt sowie den Chancen, in den USA einzureisen und die Staatsbürgerschaft zu bekommen auf der einen Seite und den Migrantengruppen, deren Ethnophaulismen am negativsten waren, fest. Sozialpsychologen vertreten sogar die Annahme, dass Ethnophaulismen genutzt werden können, um die Selbstmordsrate zwischen verschiedenen Gruppen von MigrantInnen vorherzusagen (Mullen und Smyth 2004).

## **2.2 Eine juristische Frage?**

*Hate Speech* wird in der Forschung zumeist als juristische Frage behandelt: 38 Prozent aller wissenschaftlichen Artikel mit dem Begriff *Hate Speech*, die in der Datenbank für wissenschaftliche Publikationen, dem *Web*

of Science, indexiert sind, kommen aus der Rechtswissenschaft<sup>2</sup>. Im deutschsprachigen Raum wird dieser Konfliktgegenstand hauptsächlich unter dem Begriff Volksverhetzung diskutiert und stellt dort nach §130 des deutschen Strafgesetzbuches einen Straftatbestand dar. Im deutschen Strafgesetzbuch wird Volksverhetzung wie folgt definiert (Strafgesetzbuch):

„Wer in einer Weise, die geeignet ist, den öffentlichen Frieden zu stören,

1. gegen eine nationale, rassische, religiöse oder durch ihre ethnische Herkunft bestimmte Gruppe, gegen Teile der Bevölkerung oder gegen einen Einzelnen wegen seiner Zugehörigkeit zu einer vorbezeichneten Gruppe oder zu einem Teil der Bevölkerung zum Hass aufstachelt, zu Gewalt- oder Willkürmaßnahmen auffordert oder

2. die Menschenwürde anderer dadurch angreift, dass er eine vorbezeichnete Gruppe, Teile der Bevölkerung oder einen Einzelnen wegen seiner Zugehörigkeit zu einer vorbezeichneten Gruppe oder zu einem Teil der Bevölkerung beschimpft, böswillig verächtlich macht oder verleumdet,

wird mit Freiheitsstrafe von drei Monaten bis zu fünf Jahren bestraft.

Eine Gleichsetzung von *Hate Speech* und juristischen Definitionen ist in mehrerer Hinsicht problematisch. So beschränkt sich die bewusste und intentionale kommunikative Herstellung von Minderwertigkeit von Menschen nicht darauf, was Gesetze als *Hate Speech* betrachten und/oder auf das, was „strafbar“ ist.

---

<sup>2</sup> Suche im *Web of Science* am 23.März 2017 nach dem Suchwort *Hate Speech*.

Juristische Definitionen sind Reaktionen auf gesellschaftliche Entwicklungen und laufen sozialen Entwicklungen immer hinterher, wie der Umgang mit den sogenannten Hasspostings zeigt. Juristische Definitionen können einen Beitrag zur Regulierung leisten, sind aber nicht konstitutiv für diese Phänomene.

Auch wenn die Regulierung eine Definition verlangt, folgen juristische Definitionen der Logik des Rechtssystems, das nach Motiv und TäterIn sucht (Tsesis 2013). Das erfolgt sogar im Fall von extremeren Formen wie *Dangerous Speech* (*Hate Speech* im Rahmen offener Konflikte):

„To be convicted of incitement to genocide under international law, one must be shown to have had the specific intent to commit genocide, that is ,to destroy, in whole or in part, a national, ethnical, racial, or religious group, as such.“ (Benesch 2004, S. 66)

Das gleich gilt aber auch für *Hateful Speech*, wie z.B. „Gruppen“-bezogenen Beschimpfungen, Drohungen oder Aufrufen zur Gewalt. So erklärt Tsesis (2013, S. 1170), dass „(...) the question for the Court is whether the speaker meant a threat to be outrageous or intimidating“.

Die „Motive“ und die Intentionalität der „TäterInnen“ stellt für die Einführung von Beweisen in juristischen Prozessen eine gewaltige Herausforderung dar:

„This means that a journalist who publishes or broadcasts someone else’s views, no matter how hateful these views may be, cannot be held liable (under international law, L.S.) unless the journalist can be shown to share the opinions and genocidal aspirations of the source.“ (Benesch 2004, S. 66)

Ein Fall mag die Dimensionen der Herausforderung, vor allem auch für die betroffenen Personen, zeigen: Der brasilianische Stand-up Comedian Danilo Gentili ist wiederholt durch Witze und beleidigende Kommentare

gegen Minderheiten bekannt geworden. Er wurde deswegen mehrfach kritisiert, angeklagt und sogar verurteilt. 2012 machte er via Twitter einen rassistischen Witz, in dem er Fußballspieler mit Affen gleichsetzte. Als ein schwarzer Mann sich bei ihm ebenfalls via Twitter beschwerte, antwortete Gentili über das soziale Netzwerk folgendes: „Wie viele Bananen möchtest Du, um das Ganze zu vergessen?“. Gentili blieb mit der Begründung straf-frei, er habe nicht die Intention, andere zu beleidigen, sondern zum Lachen zu bringen (Scocuglia 2014). Die Tatsache, dass die Beleidigung nicht das Endziel, sondern durchaus eine subsidiäre Aktion und daher auch intentional erfolgt, hat in diesem Fall nicht für eine strafrechtliche Konsequenz für den *Hate Speaker* gereicht.

Prinzipiell wird nur ein geringer Bereich von *Hate Speech*-Fällen durch juristische Definitionen abgedeckt. Die juristische begrenzte Reichweite bestätigt die Rolle, die dem Rechtssystem zugeschrieben wird, nämlich die **minimalen** Bedingungen für einer friedlichen Koexistenz zu garantieren (Villanueva 2002). JuristInnen können und sollen sich mit dem Problem beschäftigen und auch das Rechtssystem muss im Fall von *Hate Speech* die Vorgabe im Rahmen seiner Aufgaben bezüglich der Garantie von Grundrechten wie der Menschenwürde erfüllen.

Es gibt aber auch berechtigte Gründe, warum diese Rahmen prinzipiell begrenzt sind. Die Grenzen juristischer Definitionen liegen sowohl an den innewohnenden Begrenzungen des Rechtes als auch an der Komplexität des Phänomens. Nicht einmal in der Wissenschaft gibt es eine klare und umfassende Definition, was *Hate Speech* ist (Mendel 2012), so dass es nicht verwundert, dass sich JuristInnen seit Jahrzehnten ebenfalls darum streiten. Dies wiederum hat zu einer Begriffsinflation geführt und eine Festlegung noch zusätzlich erschwert. Zugleich wurde damit aber auch einer gesetzlichen Einschränkung der Meinungsfreiheit Tür und Tor geöffnet.

Die Entstehung des juristischen Tatbestandes der Volksverhetzung zeugt davon. Nach Goldberg (2015) sind europäische und deutsche Gesetze ursprünglich im 19. Jahrhundert im Zuge repressiver Kampagnen gegen die politische Linke entstanden:

„Modern hate-speech law began as reactionary, counterrevolutionary politics. There had been earlier laws limiting demeaning, hateful speech. Prussia’s 1794 law code, the Allgemeines Landrecht (ALR), banned expressions of incitement against the state and against certain religious groups, as well as speech that dishonored individuals. Art. 151, for example, punished, with a prison sentence of up to two years, speech that caused the ‚disapproval (Tadel) or ridicule (Verspottung) of the country’s laws‘ and ‚citizen discontent with the government.‘ In a similar vein, Art. 214 criminalized speech ‚insulting‘ and ‚dishonoring‘ a state-approved ‚religious society.‘“ (Goldberg 2015, S. 483)

Solche Gesetze erlebten erst nach 1945 einen Paradigmenwechsel hin zum Schutz der Menschenwürde und Minderheiten gegen Vorurteile und Gewalt:

„Germany, like many other European countries, bans speech that incites hatred (Volksverhetzung) against people on the basis of race, ethnic origin, religion, and nationality. Even Holocaust denial is illegal in certain circumstances. Such laws reflect, more broadly, a global “human rights revolution” since 1945 consisting of international treaties (e.g., the United Nation’s 1948 Universal Declaration of Human Rights) and monitoring organizations promoting the right to personal dignity and freedom from discrimination.“ (Goldberg 2015, S. 480)

Der historische Wandel des Begriffs „Volksverhetzung“ zeugt von der Abhängigkeit der Gesetzgebung von politischen Konjunkturen. Das ist ein weiterer Grund, warum *Hate Speech* nicht auf eine juristische Frage reduziert werden soll.

Juristische Lösungen funktionieren zudem nur, wenn *Hate Speech* nicht vom Staat selbst produziert oder unterstützt wird, wie dies im Fall des Holocausts und Ruandas geschehen war (Benesch 2004; Frère 2010).

*Hate Speech*-Gesetze können noch heute verwendet werden, um Meinungsfreiheit einzugrenzen. Auch kann sich deren Auslegung gegen die durch *Hate Speech* zu schützenden Gruppen wenden. So warnt Benesch (2014c) am Beispiel der Roma in Ungarn, dass *Hate Speech*-Gesetze auch benutzt werden können, um Minderheiten anzugreifen, statt sie zu beschützen:

„Most existing hate speech laws – including international, regional and national ones – are dangerously vague, in ways that are often used to restrict the freedom of speech of minorities, including preventing them from expressing legitimate grievances.“ (Benesch 2014c, S. 22).

Gagliardone et al. (2015, S. 31) zeigen ebenfalls, dass *Hate Speech*-Gesetze zur Zeit der Apartheid in Südafrika verwendet wurden, um KritikerInnen des System zu kriminalisieren. Das Potential *Hate Speech*-Gesetze gegen KritikerInnen von Rassismus einzusetzen, zeigte sich auch in Deutschland, als ein Anwalt den Schauspieler Till Schweiger wegen Volksverhetzung anzeigte. Schweiger hatte in einer Talkshow im Fernsehen seinen Unmut über rechte Demonstranten vor einem Flüchtlingsheim im sächsischen Freital geäußert (Focus 2015).

So vertritt die Jura-Professorin Mari Matsuda, dass *Racist Hate Messages* verboten werden sollen, wenn (und nur wenn):

1. The message is of racial inferiority
2. The message is directed against a historically oppressed group

3. The message is persecutorial, hateful and degrading (Matsuda 1989, S. 2357).

Wie Matsuda selbst erkennt, reicht ihre juristische Definition maximal für die Bestrafung der offensichtlichsten Formen von *Hate Speech* aus, d.h. Beschimpfungen von Menschen, Aufrufen zu Straftaten und Drohungen aufgrund einer Kategorie (Hautfarbe, Geschlecht, sexuelle Orientierung u.a.).

Die Beschränkung auf eine minimale Ebene hat wiederum außer-juristische, soziale und politische Folgen für die (De-)Legitimation von *Hate Speech*. Juristisch relevantes *Hate Speech* ist für Matsuda in erster Linie **Hateful Speech** und stellt einen offenen Ausdruck von Rassismus dar, der untypisch für Eliten ist:

„While violence and hate propaganda are officially renounced by elites, other forms of racism are not. (...) Lower- and middle-class white men might use violence against people of color, while upper-class whites might resort to private clubs or righteous indignation against ‚diversity‘ and ‚reverse discrimination‘. Institutions – government bodies, schools, corporations – also perpetuate racism through a variety of overt and covert means.“ (Matsuda 1989, S. 2334)

Beschränkt sich eine Definition auf *Hateful Speech*, wird *Hate Speech* automatisch mit Mitgliedern bestimmter sozialer Schichten verknüpft, nämlich mit „lower and middle-class white man“. Solche Definition verkennen damit aber die übergeordneten Ursachen und Quellen von *Hate Speech* selbst.

Die Verknüpfung der juristischen Dimension mit dem Phänomen selbst führt mitunter zu mehr Verwirrung als zu mehr Aufklärung. So ist ein großer Teil der Verwirrung, die die Schriftstellerin Suzanne Nossel im Begriff *Hate speech* sieht, darauf zurückzuführen:



„While only a small segment of what we call hate speech is actually criminal in the United States, the Bureau of Justice Statistics in 2012 (...) tallied nearly 300,000 real hate crimes involving violence or property damage: murders, assaults, arsons, vandalism and other acts motivated by bias against an identified group. We don't need the term 'hate speech' to describe these crimes, and by calling them hate speech, we risk implying that lesser forms of unsavory speech (such as bigotry without violent action or a highly offensive doodle) verge into criminality, as well“ (Nossel 2016, S. 2)

In der Tat braucht man den Begriff nicht, um Straftaten zu definieren, die durch Diskriminierung motiviert sind. Sie können zumeist den *Hate Crimes* zugeordnet werden und stellen kein *Hate Speech* dar. Nur ein geringer Anteil des Phänomens *Hate Speech* ist von juristischer Relevanz und daher *Hate Crimes*.

Last but not least sind Gesetze eher statisch (Saxer 1984) und können der Dynamik der Entwicklung von *Hate Speech* nicht folgen. Alles was den äußerst beschränkten Bereich der juristischen Definition überschreitet, muss durch den eher dynamischen und diesbezüglich effizienteren Bereich der Moral und der Ethik übernommen werden (Sponholz 2010).

Allerdings verlangt die Regulierung von medienvermittelten *Hate Speech*, sei durch die Ethik oder das Recht, allgemein Berücksichtigung und Kenntnis der Mediendynamik. So sind Ethikcodes für JournalistInnen oder für InternetnutzerInnen (Netiquette) sinnlos, wenn Hasspostings durch Software wie *Social Bots* generiert werden. An der Konfliktdynamik medienvermittelter Formen von *Hate Speech* kommt also keine Form der Regulierung vorbei.

Die Bekämpfung von *Hate Speech* erfordert holistische Ansätze (Ferguson 2016), in denen die gesetzliche Regulierung nur eine von mehreren Maßnahmen darstellt. Die nicht vorhandene Dynamik des Rechtes bei der Begleitung sozialer Entwicklungen, die Risiken für die Meinungsfreiheit und

die Möglichkeit einer Instrumentalisierung des Gesetzes gegen Minderheiten und KritikerInnen der Diskriminierung – all diese Faktoren sprechen dafür, das Phänomen *Hate Speech* nicht allein dem Rechtssystem zu überlassen und es so auf die Frage zu beschränken, ob es strafrechtlich relevant ist oder nicht. So empfiehlt der *Office of the United Nations High Commissioner for Human Rights*:

„While a legal response remains important, legislation is only part of a larger toolbox to respond to the challenges of hate speech. Any related legislation should be complemented by initiatives coming from various sectors of society geared towards a plurality of policies, practices and measures nurturing social consciousness, tolerance and understanding change and public discussion.“ (OHCHR 2012)

### **2.3 Was ist eigentlich Hate Speech?**

*Hate speech* wird hier als öffentliche Kommunikation bewusster und/oder intentionaler Botschaften mit diskriminierenden Inhalten verstanden. Damit handelt es sich weder um eine Frage der Sprache noch um ein des Hasses.

Im Kern handelt es sich bei *Hate Speech* um eine Form der kommunikativen Herstellung menschlicher Minderwertigkeit. Dabei werden bewusst und/oder intentional Antinomien aktiviert, in denen unterschiedliche Gruppen von Menschen als ungleichwertige und exklusive Gegensätze definiert werden. Treffen kann es viele, vor allem MigrantInnen, Frauen, MuslimInnen, JüdInnen oder Homosexuelle. Diese „Gruppen“ sind Minderheiten und/oder haben eine ungünstige politische/soziale Position im gesellschaftlichen Machtgefüge inne.

Auch wenn das Phänomen bereits während der Zeit des Nationalsozialismus beobachtet werden konnte (Tsesis 2002; Leader Maynard und Benech 2016; Yanagizawa-Drott 2014), ist der Ausdruck *Hate Speech* relativ

neu. Benesch (2014c) stellte fest, dass der Ausdruck in den großen Bucharchiven von Google erst in den 80er Jahren auftaucht und sich seitdem rasant ausgeweitet hat. Eine Folge dieser Entwicklung ist es, dass es weder in der Wissenschaft noch im Gesetz eine einheitliche Definition von *Hate Speech* gibt (Mendel 2012).

Für den deutschen Linguisten Jörg Meibauer (2013, S. 2) ist *Hate Speech* zum Beispiel eine Frage der Sprache, präziser gesagt,

„(...) der sprachliche Ausdruck von Hass gegen Personen oder Gruppen (...), insbesondere die Verwendung von Ausdrücken, die der Herabsetzung und Verunglimpfung von Bevölkerungsgruppen dienen.“ (Meibauer 2013, 2, Herv. L.S.).

Juristische Quellen sehen das jedoch anders. In der Empfehlung 97 (20) des Ministerkomitees des Europarats wird Hassrede wie folgt definiert:

„(...) all **forms of expression** which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin.“ (2009, 3, Herv. L.S.)

Der US-amerikanische Jurist John T Nockleby versteht *Hate Speech* als

„(...) **communication** of animosity or disparagement of an individual or a group on account of a group characteristic such as race, color, national origin, sex, disability, religion, or sexual orientation.“ (Nockleby 2000, S. 1277)

Nach der General Recommendation 35 aus dem Jahr 2013 der Internationalen Anti-Rassismus-Konvention (CERD) umfasst rassistisches *Hate Speech* folgende Sachverhalte:

- (1) Verbreitung von Botschaften der rassistischen Überlegenheit

- (2) Anstiftung zur rassistischen Diskriminierung
- (3) Drohungen oder Anstiftung zur Gewalt gegen Personen oder Gruppen aufgrund der Hautfarbe, des Phänotyps, der Abstammung, der nationalen oder ethnischen Herkunft
- (4) Beschimpfungen/Beleidigungen/Verspottungen/Verleumdungen aus den gleichen Gründen wie 3)
- (5) Rechtfertigung oder Leugnung von Genoziden und Verbrechen gegen die Menschlichkeit

Nicht nur die Definitionen sind uneinheitlich, der Ausdruck selbst sorgt für Verwirrung. Anders als der Name nahelegt, ist *Hate Speech* weder notwendigerweise von Hass getrieben noch beschränkt es sich auf sprachliche Äußerungen. Aus diesem Grund ist die deutsche Übersetzung „Hassrede“ irreführend bzw. falsch. In der deutschen Alltagssprache kommt das Wort „Hetze“ (Duden: „feindliche Stimmungsmache“) dem Phänomen viel näher. Allerdings hat das Wort, wie die juristische Definition von Volksverhetzung („Störung des öffentlichen Friedens“) den „Beigeschmack“, *Hate Speech* an seinen gesellschaftlichen Folgen zu messen. Wie gleich gezeigt werden soll, können die Folgen von *Hate Speech* jedoch vielfältiger sein, insbesondere für die betroffenen Personen selbst.

Um die Probleme zu umgehen, die diese Polysemie mit sich bringt, versuchten *Hate Speech*-ForscherInnen diesem Phänomen einen neuen Namen zu geben: *Words That Wound* (Delgado und Stefancic 2004), *Assaultive Speech* (Matsuda 1993), *Inflammatory Speech* (Tsesis 2013), *Excitable Speech* (Butler 2013), *Harmful Speech* (Leets 2002, S. 353), *Extreme Speech* (Hare und Weinstein 2009) u.v.a. All diese Begriffe bringen jeweils Vor- und Nachteile mit sich. Auf der einen Seite betonen sie und tragen dadurch dazu bei, eine bestimmte Facette von *Hate Speech* ausführlicher theoretisch und empirisch zu analysieren. Auf der anderen Seite können sie zu einer Begriffsinflation führen, die verwirrt und eine Erkenntniskumulation und damit die Theoriebildung erschwert.

Die Versuche das Label *Hate Speech* durch ein anders zu ersetzen, blieben jedoch bisher relativ erfolglos, denn die Alltagssprache scheint sich hier durchgesetzt zu haben. Zudem beschreibt der Begriff ein Phänomen, das im gesellschaftlichen Alltag existent ist: Die bewusste und oft intentionale Erniedrigung von Menschen durch Botschaften, die Aufrufe, die Rechtfertigung und/oder die Verharmlosung von Gewalt aufgrund einer Kategorie (Geschlecht, Phänotyp, Religion oder sexueller Orientierung). Was aber macht das Phänomen aus? Welche sind die Wesensmerkmale dieses Phänomen? Was gehört nicht dazu?

### 2.3.1 Der Hass

Die Bezeichnung *Hate Speech* stellt selbst Ursache für dessen definitorische Unschärfe dar. Der Ausdruck Hass sorgt dafür, dass *Hate Speech* als ein emotionales Phänomen behandelt wird. So benutzt Post (2009, S. 123) die Definition von *Hate* als „(...) an emotion of extreme dislike or aversion“, um *Hate Speech* als Ausdruck von Hass zu definieren. Waltmann und Haas (2011) wiederum setzen *Hate* mit extrem negativen Gefühle und *Beliefs* gleich: „Because hatred is an emotion **that one** feels for a social group, hatred, unlike anger, need not be the result of personal injury or insult and is more likely to prompt **deliberative** action“, so Waltman und Haas (2011, S. 33).

Die Definition von *Hate Speech* als Ausdruck von Hass hat zwei Folgen: Zum einen ist Hass etwas, was in einzelnen Individuen entsteht. Folgt man dem, würde die Kommunikation menschenverachtender Ideologien wie der des Nationalsozialismus zu einer individuellen Angelegenheit.

Die Definition von *Hate Speech* als eine Frage des Hasses führt zudem dazu, dass dieses als eine affektuelle Handlung (Weber 1988, S. 12) und damit als irrationales Phänomen begriffen wird. Daraus folgt, dass ein rationalen Umgang mit *Hate Speech* als Lösung für das Problem betrachtet wird (Lanning 2012).

Die Idee von *Hate Speech* als Ausdruck von Hass verschleiert auch eine grundlegende Eigenschaft von Konflikten – den Kampf um (begrenzte) Ressourcen (Coser 1964; Cobb und Elder 1972).

Die Gleichsetzung dieser Kommunikationsform mit einem Gefühlsausdruck erschwert zudem die empirische Erfassung:

„Das Erkennen von Äußerungen, die als ‚Hassrede‘ qualifiziert werden können, ist nicht zuletzt deshalb ein schwieriges Unterfangen, weil im Rahmen einer Hassrede nicht zwangsläufig ‚Hass‘ oder Emotionen zum Ausdruck gebracht werden. (Weber 2009, S. 6)

Dass *Hate Speech* keine Frage von Emotionen darstellt, bezeugen Fälle der Anwendung dieser Art von Kommunikation als politische Propaganda, wie in den Wahlen in Kenia 2008 geschehen (Gichuhi Kimotho und Nyaga 2016; Gustafsson 2016) oder die weitgehende Argumentationsprozesse in Texten von *Online Hate Groups* wie *Stormfront* (Meddaugh und Kay 2009).

Das Problem scheint aber nicht nur an der *Hate Speech*-Debatte selbst zu liegen, sondern auch an einem pauschalen Umgang mit dem Begriff „Rationalität“ und damit „Irrationalität“, was in der Alltagssprache zu einem Synonym für „emotional“ geworden ist. Dabei wird erstens übersehen, dass „Rationalität“ unterschiedliche Bedeutungen aufweisen kann und zweitens, dass je nach dem Rationalitätstyp „irrational“ nichts mit Emotionen oder Gefühle zu tun haben muss.

Max Weber (1988, 11ff.) unterscheidet diesbezüglich zwischen Zweckrationalität und Wertrationalität. Bei der ersteren gelten Handlungen als rational, wenn diese mit bewussten, zu erreichenden Ziele verknüpft werden. Eine Spielart dieser Rationalitätsform bietet die *Rational Choice Theory*, nach der „(...) an individual is rational if, and perhaps only if, he selects the

most efficient means to achieve his goals“ (Poole 1992, S. 200). Wertrational handelt wiederum derjenige, der seine Handlungen wertkonform gestaltet bzw. dies versucht.

In Anlehnung an Weber sieht Habermas (1981a) Rationalität als das Ergebnis eines historischen Modernisierungsprozesses, der durch Rationalisierung und Entmythologisierung gekennzeichnet ist. In vormodernen Gesellschaften hatte Gott das Sagen, politische Autorität überschneidet sich mit familiären/Clanartigen Hierarchien (Habermas 1981b, S. 458). Mit der Entmythologisierung wird Gott als Ordnungsgeber und -garant vom Thron gestoßen. Die damit einhergehende Rationalisierung führt ihrerseits dazu, dass Handlungen begründet werden müssen (Habermas 1981a). In modernen Gesellschaften funktioniert Rationalität deswegen als Legitimationsquelle.

In diesem Kontext wurde Rationalität zum Wert, d.h. zum Ziel- und Bewertungsmaßstab, so dass „irrational“ zu sein mit unvernünftig verbunden wird. Damit werden Handelnde und Handlung abgewertet (Gosepath 2010, S. 2206; Kuhlmann 1999; Phillips 1996).

Unterschiedliche Formen von Rationalität können aber auch unterschiedlich normativ bewertet werden. Zweckrationalität mündet für Habermas in strategischem Handeln. Dieses unterscheidet sich von kommunikativen Handlungen dadurch, dass es allein und ausschließlich Ziel-, und nicht Verständigungsorientiert ist (Habermas 1981a, S. 29).

Kommunikative Handlungen unterscheiden sich von strategischen Handlungen dadurch, dass erstere auf eine Verständigung zwischen den KontrahentInnen abzielen. Für Habermas handelt derjenige rational, der unter geeigneten Umständen selbst in der Lage ist, Gründe anzuführen (Habermas 1981a, S. 31).

Der Begriff des kommunikativen Handelns, so Habermas (1981a, S. 128), bezieht sich auf die Interaktion von mindestens zwei sprach- und handlungsfähigen Subjekten, die eine Verständigung über die Handlungssituation suchen, um ihre Handlungspläne und damit ihre Handlungen einvernehmlich zu koordinieren:

„Beim kommunikativen Handeln wird sogar der Ausgang der Interaktion selbst davon abhängig gemacht, ob sich die Beteiligten untereinander auf eine intersubjektiv gültige Beurteilung ihrer Weltbezüge einigen können. Diesem Handlungsmodell zufolge kann eine Interaktion nur in der Weise gelingen, dass die Beteiligten miteinander zu einem Konsens gelangen.“ (Habermas 1981a, S. 157)

In diesem Zusammenhang definiert Gosepath (2010, S. 2205) Rationalität die Fähigkeit von Personen, Verfahren des Begründens oder Rechtfertigens zu entwickeln, ihnen zu folgen und über sie verfügen zu können. Gegenstand von Rationalität können Personen sowie ihre Handlungen, Meinungen, Wünsche und Normen sein. Als allgemeiner Begriff schlägt der Autor deswegen vor, Rationalität mit Wohlbegründetheit gleichzusetzen. Damit ist Begründungsrationalität (Steenbergen et al. 2003) gemeint.

Webersche Rationalität lässt sich aber nicht auf die Begründung – die Nennung der Gründe – reduzieren. Mehr als das: Nicht jede Begründung wird als rational betrachtet. So unterscheidet Max Weber zwischen folgenden Bestimmungsgründen sozialen Handelns:

- (1) Zweckrational: „durch Erwartung des Verhaltens von Gegenständen der Außenwelt und von anderen Menschen und unter Benutzung dieser Erwartungen als ‚Bedingungen‘ oder als ‚Mittel‘ für rational, als Erfolg, erstrebte und abgewogene eigne Zwecke“
- (2) Wertrational: „durch bewussten Glauben an den - ethischen, ästhetischen, religiösen oder wie immer sonst zu deutenden –



unbedingten Eigenwert eines bestimmten Sachverhalts rein als solchen und unabhängig vom Erfolg“

- (3) Affektuell: „insbesondere emotional: durch aktuelle Affekte und Gefühlslagen“
- (4) Traditional: „durch eingelebte Gewohnheiten“ (Weber 1988, S. 12)

Betrachtet man also die Begrifflichkeit von Weber, wäre irrational das, was traditionell oder affektuell, nicht zielführend ist oder, was gegen Werte verstößt. In diesem Sinn kann der Holocaust bzw. der Massenermord an den Juden, Sinti und Roma sowie an anderen Minderheiten im Rahmen des Nationalsozialismus durchaus als „rational“ angesehen werden (vgl. Horkheimer und Adorno 2016). Solange das Vorgehen einen Zweck erfüllte, war es rational:

„Das Jahrhundert der Rationalität war auch das Jahrhundert hocheffizienter Maschinerien zur Vernichtung von Millionen Menschen. Die Reduzierung von Vernunft auf das Element der Zweckrationalität, die nach ihren Zwecken nicht mehr fragt, sondern nur noch die Mittel optimiert, fand hier zu einer Konsequenz, die die positive Kennzeichnung des Menschen als „vernunftbegabt“ endgültig ad absurdum zu führen droht.“ (Kuhlmann 1999, S. 15)

Hier wird klar, wie relevant die Begriffsbestimmung für die Anwendung von Rationalität als Bewertungsmaßstab ist:

„Das Vertrauen in eine rationale Gestaltung der Welt ist tief verunsichert. Was mit dem Schrecken über den ‚Terror der Vernunft‘ im Zuge der französischen Revolution begann, findet im Erschrecken über den rational gesteuerten Völkermord, über die Möglichkeiten der Biotechnologie und in Ansehung von Massenvernichtungswaffen seine Fortsetzung. Damit ist nicht das Projekt der Aufklärung als solches gescheitert, wohl aber in der Form eines **naivem Vernunftoptimismus** widerlegt.“ (Depenheuer 2003, S. 8, Herv. L.S., Herv. L.S.)

So wie der Holocaust kann *Hate Speech* in diesem Kontext durchaus rational sein. Es kann zudem unterschiedliche Formen von Rationalität annehmen. Die kommunikative Herstellung von Ungleichwertigkeit zwischen Menschen kann durch ausgereifte Argumentationsprozesse unterstützt werden, wie z.B. in den Diskursen von *Online Hate Groups*, und damit eine begründungsrationale Form annehmen. Es wird zudem von Rechtspopulisten im Wahlkampf bewusst, geplant und zwecksrational verwendet.

Was den Zusammenhang zwischen (Ir-)Rationalität und Emotionalität betrifft, wirft das die Frage auf, wo der „Hass“ bei *Hate Speech* verortet werden soll: An dem emotionalen Zustand der SprecherInnen? An der Botschaft selbst? Am Publikum? (Benesch 2014c) Wer hasst, wer wird gehasst? Ist dieses Gefühl Ursache oder Folge? (Whillock und Slayden 1995).

Die irreführende Bezeichnung verdeckt damit unterschiedliche Formen, Anwendungen sowie Ursachen und Folgen von *Hate Speech*. So ist ein emotionaler Zustand der SprecherInnen bei *Hateful Speech* unerlässlich, für *Hate Fomenting Speech* ist aber Hass bzw. Feindlichkeit das Ergebnis und nicht der Ausgangspunkt.

Das *Hate* in der Bezeichnung führt zur Verwirrung (Saleem et al. 2016, S. 2), besitzt aber – wie die meisten „Begriffe“ aus der Alltagssprache – nicht nur Unschärfe, sondern auch eine gewisse Berechtigung. *Hate* in diesem Fall bezieht sich auf Feindlichkeit, präziser gesagt, auf Feindlichkeit gegen Menschen aufgrund einer Kategorie (Religion, Geschlecht, Hautfarbe, sexuelle Orientierung usw.). Solche Feindlichkeit kann irrational im Sinne von emotional sein. Es kann aber auch (und sogar gleichzeitig) in einer Weltanschauung begründet sein (vgl. dazu Klein und Zick 2013).

Zusammengefasst ist nicht *Hate*, sondern Feindlichkeit für *Hate Speech* in *Othering*-Prozessen maßgebend. Wie zu zeigen sein wird, entstehen aus diesen *Othering*-Prozessen nicht nur gegensätzliche, sondern auch exklu-

sive Antinomien („Entweder wir oder sie“). Die Rolle von *Hate Speech* besteht darin, solche Antinomien zu aktivieren. Gefühle können Ursache, aber auch Ergebnis eines rational produzierten Diskurses sein.

### 2.3.2 Die Rede

Anders als der Begriff nahelegt, kann *Hate Speech* auch vollkommen auf Wörter verzichten (Gichuhi Kimotho und Nyaga 2016, S. 191). Auch darum ist die deutsche Übersetzung des Begriffes Hassrede irreführend bzw. falsch.

Das Aufstellen brennender Kreuze vor den Häusern afroamerikanischer Familien in den USA bietet ein klassisches Beispiel dafür, dass *Hate Speech* nicht darauf beschränkt werden kann, was gesagt oder geschrieben wird (Matsuda et al. 1993; Nockleby 2000). Auch in sozialen Netzwerken sind es häufiger Bilder als Wörter, die feindliche Reaktionen oder Hasspostings auslösen (Sweetser und Lariscy 2008). Deswegen besteht *Hate Speech* aus **Botschaften** und nicht aus Geschriebenen oder Gesagtem (Mendel 2012; Weber 2009; Nockleby 2006; Matsuda 1993; Butler 2013). Die Sprache ist nur ein möglicher Code.

Diese Botschaften entfalten erst in ihrem Kontext ihren Charakter als Form der kommunikativen Herstellung menschlicher Ungleichwertigkeit. Brennende Kreuze können ohne den Hintergrund der rassistischen Morde des Ku-Klux-Klans keine diesbezügliche Bedeutung annehmen. Auch deshalb ist *Hate Speech* keine Frage der Sprache, sondern eine der Kommunikation.

Während Kommunikation Austausch, Verständigung, Interaktion umfasst, ist Sprache nur ein Mittel dafür. *Hate Speech* kann sich einer Sprache, d.h. eines Zeichensystems (Lewandowski 1994; Bußmann 1990) bedienen, lässt sich jedoch nicht auf ein solches System begrenzen. Kommunikation (König und Stathi 2010, S. 47) besteht auch aus SprecherInnen, HörerInnen, Medium, Form, d.h. aus einem Kontext, der die eingespielten

Handlungszusammenhänge bildet, in denen Äußerungen oder Botschaften einen Sinn bekommen (Switalla 1973, S. 268).

Das ist besonders für Medien von Bedeutung, da diese Botschaften mit unterschiedlichen Codes und Zeichensystemen vermitteln und diese auch auf vielfältige Art und Weise einsetzen und kombinieren. So kann man Ungleichwertigkeit zwischen Menschen mit unterschiedliche Mitteln herstellen: satirische Bilder, Rumoren, *Fake News*, logischen Argumentationen u.a. (Ferguson 2016, S. 7).

## **2.4 Die Wesensmerkmale von Hate Speech**

Trotz der definitorischen Unterschiede haben die verschiedenen Ansätze gemein, dass Öffentlichkeit, Kommunikation und Diskriminierung als Wesensmerkmale von *Hate Speech* betrachtet werden (Saleem et al. 2016; Benesch 2012; Ben-David, A., & Matamoros Fernández, A 2016; Meibauer 2013; Delgado und Stefancic 2009; Matsuda 1993). Je nach Ansatz werden zudem Intentionalität (Matsuda 1993; Delgado und Stefancic 2004) und Performativität (Benesch 2014a) dazu gerechnet, wie gezeigt werden soll.

### **2.4.1 Wesensmerkmal Öffentlichkeit**

*Hate Speech* findet öffentlich statt (Waltman und Haas 2011; Delgado und Stefancic 2009; Weber 2009). In diesem Kontext definiert Unger (2013, S. 257) *Hate Speech* als „Publikationen, die eine Gruppe von Menschen aufgrund von geteilten Merkmalen herabwürdigen, oder eine glühende Rede, die auf der Straße eine aufgeregte Menge zu Hass oder sogar Gewalt gegenüber einer Minderheit aufstachelt“. Private Gespräche – egal um welchen Inhalt – sind kein Gegenstand dieser Art von Diskurs.

Etwas privat oder öffentlich zu sagen ist nicht das Gleiche. Wie im nächsten Kapitel zu zeigen sein wird, unterliegen öffentliche und publizistische Kommunikationskonflikte spezifischen normativen Ansprüchen. Das hat weitgehende Konsequenzen für den Kontroversegegenstand *Hate Speech*.

So finden Ethnophaulismen und *Racial Slurs* (vgl. dazu Hom 2008) durch öffentliche SprecherInnen seltener Zugang zur Öffentlichkeit als durch Nicht-öffentliche SprecherInnen (Vobič et al. 2013).

Das heißt konkret, dass das Verhältnis zwischen *Hate Speech* und Medien durch ein Kommunikationstabu geprägt werden kann („Darüber redet man nicht.“/„Das thematisiert man nicht.“). Ein Kommunikationstabu kann jedoch durch ein Sprachtabu umgegangen werden („Solche Wörter benutzt man nicht“). So kann man schwarze Menschen abwerten ohne das Wort „N\*\*er“ zu verwenden.

#### 2.4.2 Wesensmerkmal Diskriminierung

In der *Hate Speech*-Forschung besteht einen Konsens darin, dass der „Gruppen“-Bezug ein wesentliches Merkmal von *Hate Speech* ist (Silva, L., Mondal, M., Correa, D., Benevenuto, F., & Weber, I. 2016; Gagliardone et al. 2015; Meibauer 2013; Unger 2013; Benesch 2012; Delgado und Stefancic 2004; Matsuda 1993). Im „Gruppen“-Bezug liegt der Unterschied zwischen *Hate Speech* und anderen kommunikativen Formen von Abwertung, wie Aufrufen zu Straftaten, Beschimpfungen, Beleidigungen und Flüchen. Diese müssen nicht unbedingt gruppenbezogen sein. In diesem Sinn definiert Unger (2013, S. 259) *Hate Speech* als „Äußerungen, die eine Person oder eine Gruppe **auf der Grundlage eines gemeinsamen Merkmals** degradieren“ (Herv. L.S.).

Der „Gruppen“-Bezug bei *Hate Speech* entsteht durch Kategorisierung. Eine Kategorie ist ein Attribut, wodurch Menschen klassifiziert werden. Einer Kategorie lassen sich unterschiedliche Merkmale zuordnen. Nach der Kategorie Religion z.B. können Menschen als MuslimInnen, ChristInnen, HinduistInnen usw. kategorisiert werden.

Kategorien kennzeichnen eine Anzahl von Menschen als soziale **Einheit**. Entscheidend für soziale Einheiten ist, dass derartige gedankliche Personengebilde in der sozialen Wirklichkeit Einzelmenschen mit spezifischen

Eigenschaften sind, die nicht miteinander in Kontakt stehen (Markefka 1990).

Anders als soziale Kategorien stehen soziale Gruppen in Kontakt miteinander, empfinden sich als Einheit und treten als solche äußerlich in Erscheinung (Markefka 1990). So stellen Katholiken eine soziale Kategorie dar, die Gemeinde der Peter-und-Pauls-Kirche bildet aber eine soziale Gruppe.

Im Kontext von Diskriminierung sorgt die Kategorisierung dafür, dass Menschen mit einem gewissen Merkmal in ein soziales Gebilde, in eine „Gruppe“ verwandelt werden. Die Herstellung einer „Gruppe“ ist eine Leistung oder Performance der Kategorisierung.

Dies erfolgt z.B., wenn MuslimInnen als ein vernetztes Ganze wahrgenommen und dargestellt werden. Dadurch werden sie zu einer Gesamtheit, die kollektiv handelt und die gleichen Interessen verfolgt. Eine logische Folge davon ist die Forderung, dass die MuslimInnen sich für die Handlung von tatsächlichen Gruppen wie Al-Qaeda oder Daesh (der sich selbst als „Islamischer Staat“ bezeichnet) entschuldigen oder davon distanzieren sollen. Aus diesem Grund ist die Verwandlung einer Kategorie in eine Gruppe nicht unbedingt der Ausgangspunkt, sondern das Ergebnis von Diskriminierung (Spielhaus 2010, 12f.).

Kategorisierung ist keine Exklusivität symbolischer Diskriminierung. Sie kann sowohl als ein natürlicher, kognitiver Vorgang vorkommen (Meibauer 2013, S. 3) als auch bewusst/rational verwendet werden, wie es im Fall der Wissenschaft geschieht (Thiele 2015). Kategorisierung allein ist deswegen weder für Diskriminierung noch für *Hate Speech* ausreichend.

*Hate Speech*-Botschaften gehen weiter als Kategorisierung, indem sie eine solche „Gruppe“ mit einer Bewertung versehen (Wagner 2001, S. 13). Die Kombination von Kategorisierung und Evaluation mündet schließlich in Diskriminierung. Beispiele dafür sind Rassismus, Sexismus, Ausländerfeindlichkeit, Islamophobie oder Homophobie.

Nicht jede Form von Diskriminierung stellt jedoch *Hate Speech* dar. *Hate Speech* besteht aus einer symbolischen Form von Diskriminierung. Rassismus, Sexismus, Ausländerfeindlichkeit, Islamophobie oder Homophobie unterscheiden sich von *Hate Speech* dadurch, dass sie nicht nur symbolisch, sondern auch auf materieller Weise erfolgen können.

Rassismus offenbart sich z.B. in seiner strukturellen Form der Verteilung der Ressourcen oder den Möglichkeiten des sozialen Aufstiegs. Das erfolgt, wenn Frauen und schwarze Menschen trotz gleicher Qualifikation und Leistung wie Männer bzw. weiße Menschen, nicht in Führungspositionen gelangen können (Moore 2016). Klassischer Rassismus, Islamophobie und Ausländerfeindlichkeit treten können ebenfalls in einer systemischen bzw. institutionalisierten Form in Erscheinung treten wie z.B. im Falle des *Racial Profiling* (Frost 2008, S. 554).

Solche Arten „Gruppen“-bezogener Menschenfeindlichkeit unterscheiden sich von *Hate Speech* zudem dadurch, dass sie nicht bewusst und absichtlich sein müssen, sondern auch in einer alltäglichen und naturalisierten Form vorkommen können. Das erfolgt z.B., wenn schwarze Menschen auf Fragen bei Bewerbungsinterviews antworten müssen, die für die Ausübung der Funktion keine Rolle spielen und, die andere KandidatInnen nicht beantworten müssen. Dazu gehört z.B., die Frage für schwarze Frauen in Vorstellungsgesprächen, ob sie gut tanzen können (Essed 1991, S. 142).

Diese Handlungen fallen unter Rassismus, d.h. einer Unterscheidung, Ausschließung, Beschränkung oder Bevorzugung, die auf der Hautfarbe, des Volkstums, des nationalen Ursprung oder religiösen Intoleranz beruht (UNESCO 1978). Keine von denen stellt jedoch *Hate Speech*.

*Hate Speech* ist eine **Kommunikationsform** „Gruppen“-bezogener Menschenfeindlichkeit. Sie ist aber nicht die einzige. Wie bereits gezeigt, werden Rassismus, Sexismus, Islamophobie, Homophobie usw. ebenfalls

unbewusst und/oder nicht-intentional kommuniziert. Sie stellen daher symbolische Diskriminierungen, aber nicht *Hate Speech* dar.

Um festzustellen, ob es sich um *Hate Speech* oder um andere Formen symbolischer Diskriminierung handelt, muss daher stets die Kommunikationssituation berücksichtigt werden.

#### 2.4.3 Wesensmerkmal Kommunikation

*Hate speech* wird oft auf ein Sprachsystem reduziert, wie bei Meibauers Definition von Hassrede als „der **sprachliche Ausdruck von Hass**“ (Meibauer 2013, 3, Herv. L.S.). Wichtiger ist jedoch das, was man mit der Sprache macht, und das verwirklicht sich durch Kommunikation. Die Sprache selbst ist das Zeichensystem (Lewandowski 1994), dessen man sich bedient, ein Sprechakt dagegen die eigentliche Handlung, die man mit diesem Zeichensystem vollzieht (Switalla 1973, S. 268). Diese Handlung wird in einem Zusammenhang vollzogen, zu dem SprecherIn, HörerIn, Medium usw. gehören. Erst dadurch bekommt Sprache einen Sinn (Austin 1971). Dieser Kontext stellt die Kommunikationssituation dar. In diesem Sinn ist die Sprechakttheorie eigentlich nicht nur eine Theorie der Sprache, sondern auch der Kommunikation.

Der Philosoph John Austin und sein späterer Schüler John Searle haben einem sogenannten *linguistic turn* initiiert, indem sie den Fokus von der Repräsentationsfunktion der Sprache auf die Sprachhandlung verschoben haben (Koch 2010, S. 11). Das, was man sagt, steht nicht nur stellvertretend für etwas, sondern schafft selbst Tatsachen. Dies geschieht z.B., wenn der Pfarrer sagt: „Damit erkläre ich Sie zu Mann und Frau“. Diese Sprachhandlung generiert eine soziale Tatsache – ein Ehepaar (Searle 2008).

In diesem Kontext reduziert sich eine Äußerung nicht nur auf das, was konkret gesagt oder geschrieben wird, sondern drückt auch Absicht und Folgen aus. Diese drei Elemente werden in der Sprechakttheorie als lokutionäre, illokutionäre und perlokutionäre Akte bezeichnet. Der lokutionäre



Akt bezieht sich auf das, was gesagt oder geschrieben wurde und besteht aus einem Referenzakt und einem Prädikationsakt. In *Hate Speech*-Aussagen stellt eine „Gruppe“ oder ein zur „Gruppe“ zugehöriges Individuum den Referenzakt dar. Der Prädikationsakt bei derartigen Aussagen kann unterschiedliche Formen annehmen (Ist-Aussage, Soll-Aussage u.a.), enthält aber stets eine Bewertung.

Der illokutionäre Akt umfasst das, was man mit Sprache tun kann: behaupten, wünschen, Absichten ausdrücken, Gefühle äußern und neue Tatsachen schaffen. Dies lässt sich jedoch nicht allein aus dem, was gesagt oder geschrieben wurde, herauslesen, denn der gleiche Satz kann unterschiedliche Absichten ausdrücken (Searle 2008).

Anders formuliert: Äußerungen unterscheiden sich nach ihrer Illokutionen. Solche Illokutionen grenzen sich wiederum durch den Zweck voneinander ab, der verfolgt wird, dem psychischen Zustand, den sie ausdrücken und der Beziehung, die sie zwischen Wort und Welt aufstellen.

Repräsentative oder assertive Sprechakte beschreiben, mitteilen oder behaupten. Sie sagen, wie die Welt sich verhält. Sie drücken das, was man glaubt, aus. Direktive Sprechakte bewegen jemanden dazu, etwas zu tun. Sie wirken auf die Welt ein und drücken einen Wunsch aus. Kommissive Sprechakte wirken ebenfalls auf die Welt ein, werden aber verwendet, um sich selbst auf eine Handlung festzulegen. Das ist der Fall bei Versprechen, Vereinbarungen, Angeboten oder Drohungen. Expressive Sprechakte haben keinen Bezug zur Welt und dienen dazu, Gefühlen Ausdruck zu verleihen. Beispiele dafür sind danken, klagen oder beschimpfen. Deklarative Sprechakte schaffen neue Tatsachen, so dass sie sich sowohl auf die Welt beziehen als auch auf diese Welt einwirken. Das passiert, wenn man eine Person entlässt oder der Richter jemanden für schuldig erklärt (Searle 2008).

Tabelle 1: Typen von Illokutionen

	<b>Zweck</b>	<b>Ausrichtung</b>	<b>Psychischer Zustand</b>	<b>Beispiele</b>
<b>Repräsentativa/Assertiva</b>	sagen, wie es sich verhält	Wort auf Welt	Glaube	behaupten, mitteilen, berichten
<b>Direktiva</b>	jemanden zu einer Handlung bewegen	Welt auf Wort	Wünsche	bitten, befehlen, raten
<b>Kommissiva</b>	sich selbst auf eine Handlung festlegen	Welt auf Wort	Absicht	versprechen, vereinbaren, anbieten, drohen
<b>Expressiva</b>	Ausdruck der eigenen Gefühlslage	keine	Zustand	danken, grüßen, beglückwünschen, klagen
<b>Deklarativa</b>	mit dem Gesagten entsprechend die Welt verändern	beide	x	ernennen, entlassen, taufen

Quelle: Searle 2008

*Hate Speech* lässt sich wiederum nicht auf einen Typ von illokutionärem Akt begrenzen. So ist *Hate Speech* nicht auf eine Assoziation einer Kategorie mit einer Wertung zu reduzieren. Mit dieser Kommunikationsform kann man viel mehr tun, als eine Kategorie von Menschen zu repräsentieren.

Tabelle 2: *Hate Speech* und Sprechakte

<b>Hate Speech-Handlungen</b>	<b>Illokution</b>
Drohung	Kommissiva
Anstiftung	Direktiva
Beschimpfung	Expressiva
Verleumdung	Assertiva
Rechtfertigung	(Mehrere Möglichkeiten)
Leugnung	Repräsentativa/Assertiva
Verharmlosung	Repräsentativa/Assertiva
Verbreitung	Kein Sprechakt

Quelle: eigene Darstellung

Perlokutionäre Akte beziehen sich wiederum auf das Ziel der SprecherInnen sowie auf die Wirkung der Äußerung auf die HörerInnen. Hierbei ist es zielführend, zwischen perlokutionärem Akt und perlokutionärem Effekt zu unterscheiden, da Absicht und Wirkung einer Aussage auseinanderfallen können: So können SprecherInnen mit einer Äußerung die Absicht verbinden, jemanden „zum Lachen bringen“ (perlokutionärer Akt), während die nachfolgende Wirkung darin besteht, dass die HörerInnen beleidigt sind (perlokutionärer Effekt) (Herrmann und Kuch 2007, S. 13; Hitzler 1989).

In diesem Fall scheitert der perlokutionäre Akt (Butler 2013, S. 31; Krämer 2010, S. 36). Dies ist z.B. auch der Fall bei konflikthaltigen, provokativen Inhalten, die keine oder nur geringe Auseinandersetzungen verursachen, weil keine Reaktion erfolgt. Aber auch im gegenteiligen Fall sind Absicht und Wirkung nicht dasselbe: So kann eine Aussage eine Debatte oder einen Skandal auslösen, obwohl sie nicht als Provokation gedacht war, wie dies z.B. bei einem Fauxpas der Fall ist.

Diese theoretische Zuordnung hat sowohl in der Öffentlichkeit als auch in der Wissenschaft Folgen für das Verständnis von *Hate Speech*. Sie bietet vor allem eine theoretische Grundlage, um unterschiedliche Formen von *Hate Speech* zu identifizieren. Sie ermöglicht ebenfalls, *Hate Speech* von anderen Formen symbolischer Diskriminierung zu unterscheiden.

Auf Basis dieser theoretischen Überlegungen können drei Modelle von *Hate Speech* identifiziert werden: das lokutionäre Modell (*Hateful Speech*), das ilokutionäre Modell und das perlokutionäre Modell (*Dangerous Speech*).

#### 2.4.3.1 *Das lokutionäre Modell*

Wenn *Hate Speech* lokutionäre Sprechakte darstellt, dann kann diese Kommunikationsform an dem, was konkret gesagt oder geschrieben wird, identifiziert werden. Das Entscheidende an dem lokutionären Modell von *Hate Speech* ist die Verortung dessen an die Aussage selbst. Im Fall von *Hate Speech* wäre der verwendete Wortschatz für seine Erfassung bestimmend.

Die Tatsache, dass *Hateful Speech* einen lokutionären Akt darstellt, bedeutet nicht, dass ihm keine Intention (illokutionärer Akt) oder Folgen (perlokutionärer Akt) innewohnen, sondern allein, dass er durch dens verwendeten Wortschatz definiert wird. In Bezug auf die Illokution handelt es sich bei *Hateful Speech* häufig um affektuelle Handlungen. Solche Äußerungen drücken den Gefühlszustand der SprecherInnen aus und ähneln einer anderen Form verbaler Aggression – der Beschimpfung (vgl. dazu Havryliv 2009)).

Ein Beispiel liefert dieser Nutzer von YouTube auf seinem Kommentar zu einem Video um den Selbstmordversuch eines Flüchtlings in Wien:

„(...) Ich für meinen Teil SCHEISSE auf diesen verkommenen Wüstenaffen, der mir wieder zeigt das dieser unkulturelle Bodensatz durch Inzucht und jetzt, da ist SIE, die berühmte PSYCHE!!!!

[illegible]

*Hateful Speakers* sind häufig Nicht-öffentliche SprecherInnen (Gagliardone 2015) und besitzen oft einen niedrigen Bildungsgrad (Matsuda 1989). Dieser Typ von *Hate Speech* muss nicht medienvermittelt sein und kann in Face-to-Face Situationen vorkommen. Die Dynamik dieser Art von Hassrede kann aber durch Medien wesentlich geprägt werden, wie es im Fall von Hasspostings zu beobachten ist (Benesch 2014a; Gagliardone et al. 2015; Gichuhi Kimotho und Nyaga 2016).

Medien wie *Social Networking Sites*, erweitern die Reichweite der SprecherInnen und bringen sie in anderen Kommunikationssituationen als dies rassistische Beschimpfungen auf einem Universitätscampus darstellen (Kennedy 2002). Nicht nur das Medium Internet, sondern auch der modus operandi einzelner Medienunternehmen stellt einen wesentlichen Einfluss dar. In diesem Sinn stellen komparative Studien zu sozialen Medien fest, dass das personenbezogene Netzwerk Facebook dasjenige ist, in

<sup>3</sup> Verfügbar über: <https://www.youtube.com/watch?v=7fifS9cSmf8>, Zugang am 23.06.2017.

dem *Hateful Speech* am häufigsten erfolgt (Gichuhi Kimotho und Nyaga 2016, S. 193; Gagliardone 2015)

Die Erforschung von *Hateful Speech* hat eine lange Tradition in der Sozialpsychologie, insbesondere unter dem Begriff Ethnophaulismus (Leader et al. 2009; Birnbaum 1971; 1971; Leader et al. 2009; Roback 1979). Lexika-basierte Ansätze werden zudem derzeit intensiv bei der Untersuchung von sogenannten Hasspostings verwendet, da diese eine automatisierte Identifikation von *Hateful Speech* in großen Datenmengen ermöglichen (Lucas 2014; Warner, W., & Hirschberg, J. 2012). Da eine solche Automatisierung zu Wortschatz-Analysen beinahe zwingt (Saleem et al. 2016), reduzieren sie das Problem *Hate Speech* allein auf *Hateful Speech*. Einige dieser Studien gehen sogar schon von Beginn an davon aus, dass *Online Hate Speech* eine Frage des rhetorischen Stils und damit allein der Sprache ist (Neumann und Arendt 2016, S. 247).

Einen weitergehender Versuch *Hate Speech* zu erfassen ist es, nicht nur Wörter wie Ethnophaulismen oder *Racial Slurs*, sondern auch semantische Strukturen und Themen als Kriterium heranzuziehen (Burnap und Williams 2016; Gitari et al. 2015).

Der größte Vorteil des lokutionären Modells ist dessen einfache empirische Operationalisierung. Die Gleichsetzung eines Indikators (Wörter) mit einer Kommunikationsform (*Hate Speech*) ist ein pragmatischer Ansatz, um dieses Phänomen in größeren Datensätzen zu identifizieren. Die Anwendung der Lexika-basierten Ansätze in computergestützten Methoden zeigen sich zudem zuverlässig darin, zu identifizieren, wie oft *Derogatory Labels* verwendet werden und damit auch, welche Gruppen das Ziel von Angriffen werden.

Solche empirische Erfassungen bleiben jedoch auf der sprachlichen Ebene stehen. Sie ermöglichen es, offene, diskriminierende Sprache und bestimmte Typen von *Hate Speech* zu erfassen. Nicht berücksichtigt werden aber Nicht-lexikalisierte Formen von *Hate Speech*.

Das ist auch der Grund, warum es diesen Lexika-basierten Verfahren nicht gelingt, Botschaften die zu *Hate Speech* ansstiften, von anderen zu unterscheiden (Bartlett, J., Reffin, J., Rumball, N., & Williamson, S. 2014; Grabow und Hartleb 2013). Auch weitergehende Ausdifferenzierungen, Ursachen und Dynamiken bleiben in derartigen Studien ausgeblendet.

Lexika-basierte Ansätze gehen von der Grundannahme aus, dass Kategorisierung und Evaluation in Lexikalisierung münden. So wie sich Sprechakte aber nicht nur auf die Wörter reduzieren lassen, braucht Diskriminierung keine eigene Lexikalisierung, sondern kann in einen völligen alltags-sprachlichen, nicht a priori abwertendem Wortschatz auftreten.

Darum führt die Gleichsetzung von Ethnophaulismen (der Indikator) und *Hate Speech* (das Konstrukt) zu falschen Schlussfolgerungen:

„The meaning of speech depends on the context in which it occurs, and there are several factors that can affect or change it. Simply analyzing the words themselves does not tell the entire story. It is important to consider who is speaking, what his or her intention is, and what his or her relationship to the audience is. It is also important to consider the agency of the audience and its understanding of the message.“ (United States Holocaust Memorial Museum 2009)

Das könnte ein Grund dafür sein, dass sich *Hate Speech* in diesen Studien als sehr begrenztes Phänomen zeigt. Gagliardone (2015) und seine Forschungsgruppe analysierte mehr als eintausend Facebook-Seiten aus Äthiopien und stellen ein Anteil von nur 17 Prozent „Gruppen“-bezogener, beleidigender Kommentare fest. In Deutschland ermitteln Neumann und Arendt (2016) einen Anteil von weniger als 10 Prozent hasserfüllter Kommentaren gegen Flüchtlinge auf den Facebook-Seiten der Boulevardzeitung *Bild*. Auch auf den Facebook-Seiten rechtsextremer Parteien in Spanien umfasst diese Art der Kommentare nur einen Anteil von nicht mehr 12 Prozent (Ben-David, A., & Matamoros Fernández, A 2016).

### 2.4.3.2 *Das illokutionäre Modell*

Betrachtet man *Hate Speech* als illokutionären Akt, dann steht die Intention der SprecherInnen im Fokus. Wenn *Hate Speech* als lokutionärer Akt bzw. als *Hateful Speech* am Wortschatz zu erkennen ist, ist sein illokutionäres Pendant an der Intention der SprecherInnen festzumachen.

Für Searle (1980) ist Intention nicht die Folge, sondern die Ursache einer Handlung. Eine intentionale Handlung liegt vor, wenn der Grund eines Erfolges an der Intention des Handelnden liegt. Searles Beispiel mag den Unterschied zwischen intentionalen Handlungen und Handlungsfolgen verdeutlichen:

„Oedipus intended to marry Jocasta but when he married Jocasta he was marrying his mother. ‚Marrying his mother‘ was not part of the intentional content of the intention in action, but it happened anyhow. The action was intentional under the description ‚marrying Jacosta,‘ it was not intentional under the description ‚marrying his mother.‘“ (Searle 1980, S. 67).

Kurz formuliert: Ödipus wollte Jokasta heiraten und nahm die Tatsache, dass sie seine Mutter war, in Kauf. Deswegen handelt es sich in diesem Fall um tatsächlich vollzogenen Inzest, aber nicht um intentionalen Inzest.

An der Intentionalität liegt der Unterschied zwischen *Hate Speech* und andere Formen symbolischer Diskriminierung. Die Abwertung einer „Gruppe“ in einer Botschaft oder in einer Äußerung kann erfolgen, auch wenn dies nicht das Ziel ist. So erfolgt z.B. bei Hilfskampagnen humanitärer Organisationen mit Bildern schwarzen Kinder (Kreuzkamp 2014). Durch die Assoziierung von Armut/Not mit der Hautfarbe werden Menschen kategorisiert und ihnen eine Stelle auf einer Werteskala zugeschrieben: Schwarze Menschen sind arm und hilfsbedürftig. Es handelt sich also um diskriminierende Äußerungen. Gleichzeitig ist die Abwertung der schwarzen Menschen nicht der Grund einer solchen Kampagne. Die Diskriminierung ist in diesem Fall nicht intentional.



Anders ist dies im Fall von *Hate Speech*: Das Ziel von *Hate Speech* ist es, Handlungen auf Basis der kommunikativen Herstellung von Ungleichheit zwischen Menschen zu vollziehen, sei es behaupten, mitteilen, zu versprechen, zu drohen, zu klagen oder zu entlassen. Das Bild hilfsbedürftiger schwarzer Menschen aus einer humanitären Kampagne auf dem Plakat einer Partei auf dem der Slogan steht „Wir sind nicht das Sozialamt für die ganze Welt“ verwandelt sich in eine von Grund auf andere Botschaft, als dies bei dem Beispiel der Hilfsorganisationen der Fall ist.

Diskriminierung muss jedoch nicht unbedingt das Endziel von *Hate Speech* sein, damit es intentional ist. Es reicht aus, dass es eine subsidiäre Handlung darstellt. Zwischenhandlungen sind in diesem Sinn ebenfalls intentional:

„For example, suppose I have a prior intention to drive to my office. As I am carrying out this prior intention I might perform a series of subsidiary actions for which I need not have formed a prior intention: opening the door, starting the engine, depressing the clutch, etc. When I formed my intention to drive to the office I might not have given these subsidiary acts a thought. Yet such actions are intentional.“ (Searle 1980, S. 52)

Wenn man ins Büro kommen will, kann man dafür ein Auto nutzen. Diese Handlung ist ein (notwendiger) Schritt, d.h. eine subsidiäre Handlung (Autonutzung), um die primäre Intention zu vollziehen (ins Büro kommen) zu erreichen, d.h. ein Zweck zur Erreichung eines Zieles.

*Hate Speech* bleibt auch dann intentional, wenn eine rassistische Gesinnung nicht gegeben ist. Intentionalität liegt auch vor, wenn Rassismus als ein Schritt bzw. eine Zwischenhandlung fungiert, um eine primäre Intention zu verwirklichen. So können abwertende Botschaften gegen eine „Gruppe“ eine subsidiäre Aktion darstellen, um Bücher zu verkaufen, um Medienaufmerksamkeit zu erhalten oder um Stimmen im Wahlkampf zu erringen.

Das illokutionäre Modell reduziert das Phänomen jedoch nicht auf die Absicht der SprecherInnen. So versuchen deren VertreterInnen immer wieder, die Notwendigkeit einer Regulierung mit der Wirkung des *Assaultive Speech* auf die Betroffenen zu begründen:

„(...) hate speech is rarely an invitation to a dialogue; it is like a slap in the face. The speaker is not inviting an intellectual discussion (...) He or she is uttering a performative – a word that enacts its own message without the need of decoding.“ (Delgado und Stefancic 2004, S. 206).

Im illokutionären Modell handelt es sich um *Words that Wound*, die genauso verletzend sein können, wie physische Gewalt. Anders formuliert: *Hate Speech* wird auch als perlokutionärer Akt verstanden. In diesem Modell wird aber die Performativität selbst nicht untersucht, sondern als Grundannahme vorausgesetzt.

SprecherInnen, die *Hate Speech* intentional verwenden, können sowohl öffentliche als auch Nicht-öffentliche Personen sein. Dieser Art von *Hate Speech* verlangt nicht unbedingt nach einer medialen Vermittlung und kann ebenfalls in Face-to-Face-Kommunikation vorkommen.

Das Wesensmerkmal der Intentionalität ist daher für das illokutionäre Modell nicht nur entscheidend um *Hate Speech* zu definieren, sondern auch um dieses gesetzlich zu regulieren. Die Intentionalität liefert JuristInnen das „Motiv“ der „TäterInnen“.

Um diese Position zu verstehen, ist es unerlässlich, den Kontext, in dem dieses *Hate Speech*-Modell entstand, zu berücksichtigen. Das illokutionäre Modell wurde von US-amerikanischen JuristInnen Ende der 80er Jahre und zu Beginn 90er Jahre im Rahmen der sogenannten *Critical Racial Theory* entwickelt.

Es entstand als Reaktion auf die Verbreitung einer Nicht-medialen Form, des *Campus Hate Speech*. Insbesondere in den 90er Jahren erlebten US-

amerikanische Universitäten eine Epidemie rassistischen Anfeindungen, bei denen es StudentInnen aber auch ProfessorInnen traf (Matsuda et al. 1993; Matsuda 1989). Diese rassistischen Fälle beschränkten sich jedoch nicht nur auf den Campus, sondern folgten dem gleichen Muster wie Angriffe auf Afroamerikaner und betrafen auch verschiedene Bereiche des sozialen Lebens. Ku-Klux-Klan-Symbole sorgten in manchen Wohngegenden dafür, dass afro-amerikanische Familien abgeschreckt wurden und aufgaben, sich Wohnungen in diesen Stadtteilen zu suchen. Cafégäste bekamen Karten mit „Schöne Grüße von der KKK“, sobald sie an einem Tisch in bestimmten Lokalen saßen, Kommilitonen wurden auf dem Campus als N\*\*\*er oder als F\*\*\*ot öffentlich geschimpft, Bürotüren von ProfessorInnen wurden mit den Initialen des Ku-Klux-Klans beschmiert, asiatische StudentInnen öffentlich beschimpft (Delgado und Stefancic 2004, S. 53; Matsuda 1989, S. 2333).

Angesichts dieser Lage suchten Jura-ProfessorInnen, die selbst auch der Zielgruppe dieser Angriffe angehörten, nach Möglichkeiten, solche Beschimpfungen und Androhungen gesetzlich zu unterbinden. So begann die Forschungstradition des *Campus Hate Speech*.

Das illokutionäre Modell war schon vielfach und aus unterschiedlichen Gründen Gegenstand von Kritik. ForscherInnen wie Matsuda (1989) werden dafür kritisiert, dass sie versuchen, Alltagserfahrungen in die wissenschaftliche Forschung zu integrieren. Sie greifen auf ihren eigenen Erfahrungen mit *Hate Speech* zurück, was ihre Arbeit Kritikanfällig machte (Sandmann 1994, S. 250). Ein weiterer Kritikpunkt betrifft die Grundannahme einer Performativität.

So kritisiert Sandmann (1994), dass „not all recipients of such speech will respond in the same way. She (*Matsuda*) offers little evidence that even the majority of recipients will respond to hate speech in the same way“ (Anm. L.S.). Eine neue Formulierung dieser Kritik findet sich bei Butler (2013),

wenn sie in Frage stellt, ob illokutionäre Sprechakte (Absichten oder Intentionen) immer erfolgreich sind, wie Matsuda unterstellt. Wie bereits in der Sprechakttheorie nachgewiesen, können Sprechakte samt Absicht scheitern: Ein Witz kann z.B. fehlgehen, so dass keiner lacht. Daraus kann man ableiten, dass die Performance der Verletzung durch *Hate Speech* ebenfalls versagen kann. Anders als die *Critical Racial Theorists* um Matsuda behaupten, könnte es sein, dass verletzende Wörter nicht verletzen<sup>4</sup>.

Das Matsuda-Modell wird ebenfalls wegen dessen Einstellung zu den Machtverhältnissen zwischen den Gruppen, die angefeindet werden, und denjenigen der *Hate Speaker*, kritisiert. Matsuda sagt ausdrücklich, dass das Ziel nur eine „historically oppressed group“ sein kann:

„Expressions of hatred, revulsion, and anger directed against dominant-group members by subordinated-group members not criminalized by the definition of racist hate messages used Malcolm X's "white devil" statements - which he later retracted are an example. Some would find this troublesome, arguing that any attack on any person's ethnicity is harmful. The harm and hurt is there, but it is of a different degree. Because the attack is not tied to the perpetuation of racist vertical relationships, it is not the paradigm worst example of hate propaganda.“ (Matsuda 1989, S. 2361)

Dieser Punkt ist jedoch in der *Hate Speech*-Forschung umstritten. Für Meibauer (2013) z.B. ist diese Kommunikationsform von Diskriminierung nicht an Machtpositionen gebunden. Potentiell kann jede Gruppe zum Ziel von *Hate Speech* werden, auch z.B. die Fans eines Fußballclubs:

„(...) wir finden staatliche Hassrede, Hassrede von Minderheiten gegen eine unterdrückende Majorität, Hassrede von Minderheiten

---

<sup>4</sup> Es sind aber mittlerweile eine Reihe von empirischen Studien entstanden, die zeigen, dass dies tatsächlich erfolgt (siehe hierzu Kap. 2.1.).

gegen eine unterdrückende Majorität, aber auch Hassrede von Minoritäten untereinander, zum Beispiel von Marokkanern gegen über Türken und gekehrt in Deutschland.“ (Meibauer 2013, S. 2)

Damit wäre also *Hate Speech* beispielweise nicht nur islamophobisch, sondern auch das Werk von islamistischen Predigern, auch wenn sich Ursachen, Dynamiken und Umgang mit beiden Formen von *Hate Speech* wesentlich voneinander unterscheiden (Frost 2008, 555f.).

Die Abkopplung von Machtposition und *Hate Speech* ist mit dem Matsuda-Modell nicht kompatibel (vgl. auch Schwartzman 2002). Wie Delgado und Stefancic argumentieren:

„Because whites in this society enjoy more power and influence than blacks, Latinos, Asians and Indians, they are more apt to use hate speech to enforce the status quo and protect the privileges of whiteness.“ (Delgado und Stefancic 2004, S. 208)

Diese Position wird durch die Tatsache verstärkt, dass dieses Modell sich hauptsächlich auf rassistischen *Hate Speech* bezieht. Rassismus ist schon definitorisch von Macht nicht zu trennen. Nach der Erklärung über „Rassen“ und rassistische Vorurteile der UNESCO besteht Rassismus bei:

„(...) jeder auf der ‚Rasse‘, der Hautfarbe, dem Volkstum, dem nationalen Ursprung oder der von rassistischen Überlegungen getragenen religiösen Intoleranz beruhende Unterscheidung, Ausschließung, Beschränkung oder Bevorzugung, welche die souveräne Gleichheit der Staaten und das Recht der Völker auf Selbstbestimmung beseitigt oder gefährdet oder welche das Recht jedes Menschen und jeder Gruppe auf volle Entfaltung in willkürlicher und diskriminierender Weise begrenzt.“ (UNESCO 1978)

Um das Recht der Völker auf Selbstbestimmung zu beseitigen oder zu gefährden oder das Recht von Menschen auf volle Entfaltung zu begrenzen, muss man über die Mittel verfügen, um das zu vollziehen. Darum ist

Rassismus nicht von Macht zu trennen. Der Grundbegriff von Rassismus spiegelt sich im Verständnis von *Hate Speech* in der *Critical Race Theory* und dem illokutionären Modell wieder:

„Purely racial insults and name-calling (honky, cracker, whitey) directed against white do not evoke and call up a specific oppressive history for the white majority. (...) On the other hand, words such as nigger, wop, spic, chink or kike do carry a historical message that often multiplies their impact.“ (Delgado und Stefancic 2004, S. 176)

Matsuda geht klar und eindeutig von einem strukturellen Begriff von Rassismus aus. Für sie umfasst Rassismus eine Ideologie der rassischen Vorherrschaft und den Mechanismen, um die Opfergruppen in einer unterlegenen Position zu halten. Anders als ihre KritikerInnen behaupten (Sandmann 1994, S. 250), unterscheidet die Juristin durchaus zwischen Rassismus und rassistischem *Hate Speech*. Dieses wird nur als eine der möglichen „implements of racism“ gesehen. Sie systematisiert sogar unterschiedliche Formen anhand der SprecherInnen, der Formen der Diskriminierung und dem Ort (Matsuda 1989, S. 2332).

Daraus entsteht aber ein innewohnender Widerspruch dieses Modells: Ein übergeordnetes Phänomen wie Rassismus in Einklang mit einer juristischen Definition zu bringen. Obwohl diese Forschungstradition erkennt, dass *Hate Speech* nicht auf Individuen zurückzuführen ist, führt der Fokus von *Hate Speech* als eine juristische Frage zwangsläufig dazu, dass die Ursache bzw. die Quelle von *Hate Speech* bei einzelnen SprecherInnen verortet wird. Wenn *Hate Speech* eine Straftat ist, dann muss es TäterIn und Motiv (dessen Intention) geben (vgl. hierzu Kap. 2.2).

#### 2.4.3.3 Das perlokutionäre Modell

Versteht man *Hate Speech* als lokutionären Akt, dann kann es daran festgemacht werden, was geschrieben oder gesagt wurde. Das illokutionäre

Modell verortet es wiederum an der Intention der SprecherInnen. Als perlokutionärer Akt liegt *Hate Speech* wiederum an den hervorgerufenen Folgen und Effekten.

Die Haupteigenschaft dieser Spielart von *Hate Speech* ist der Zusammenhang mit offener Gewalt. Es kommt vor „when an act of speech has a reasonable chance of catalyzing or amplifying violence by one group against another, given the circumstances in which it was made or disseminated“ (Benesch 2012, S. 2). Darum nennt Benesch es *Dangerous Speech*. Dies kann bei *Hate Speech* im Vorfeld oder im Rahmen offener Konflikte der Fall sein, wie z.B. einem Genozid.

*Dangerous Speakers* sind in erster Linie Personen des öffentlichen Lebens, einmal, dass die Anwendung von Gewalt Legitimation verlangt:

„Speech in an essential tool for this conditioning, of course, as it is for any collective human effort. Periods preceding genocide, massacres or ethnic cleansing typically see inflammatory public speech from an array of influential sources – politicians and comedians, athletes and bartenders.“ (Benesch 2014b, S. 6).

Weil Verbreitung eine wesentliche Rolle bei *Dangerous Speech* spielt, verlangt dieser Art von *Hate Speech* mediale Vermittlung. Diskriminierung, Intentionalität, öffentliche SprecherInnen und Medien sind jedoch noch nicht genug, damit *Hate Speech* offene Gewalt entfacht. In diesem Zusammenhang schlägt Benesch (2012) fünf Variablen vor, anhand derer man idealtypisch *Dangerous Speech Acts* identifizieren kann:

### **(1) Die SprecherInnen, deren Profil und Popularität**

*Hate Speech* hat höhere Chancen erfolgreich anzustiften, wenn die SprecherInnen einen schon vorhandenen Einfluss oder Prestige beim Publikum haben.

### **(2) Das Publikum und sein emotionaler Zustand**

Ein ängstliches Publikum ist anfälliger für Anstiftungen zur Gewalt.

**(3) Der Sprechakt, d.h. der Inhalt von *Hate Speech* als Aufruf zur Handlung**

Anstiftung folgt gewissen rhetorischen Mustern und soll durch das Publikum als Aufruf zur Gewalt verstanden werden.

**(4) Der historische und soziale Kontext**

Dazu gehört ein langer Kampf von Gruppen um Ressourcen, mangelnde soziale und politische Mechanismen, um Missstände zu lösen oder frühere Erfahrungen mit Gewalt (besonders, wenn sie auf anstiftende Rede folgten).

**(5) Die Medien bzw. die verwendeten Mittel, um solche Botschaften zu verbreiten**

Ein Beispiel dafür liefern Mediensysteme, in denen tatsächlich oder beinahe ein einzelnes Medium als einziger Quelle von Nachrichten existiert (Frère 2010; Benesch 2014b; 2011, S. 391).

Nicht alle Bedingungen müssen gegeben sein, so dass z.B. eine geringere Popularität der SprecherInnen durch eine weitgehende Verbreitung kompensiert werden kann. So können Gerüchte über Koranverbrennungen in Pakistan und Afghanistan Gewalt gegen Personen oder Gruppen auslösen, auch wenn die Quelle nicht glaubwürdig ist (Leader Maynard und Benesch 2016, S. 79). Ebenso können Äußerungen prominenter SprecherInnen Gewalt auslösen, auch wenn sie nur durch ein unbedeutendes Medium verbreitet werden.



Diese Unterscheidung ist besonders wichtig, weil es keinen zwingenden Zusammenhang zwischen *Hate Speech* und offener, physischer Gewalt gibt (Leader Maynard und Benesch 2016, S. 71).

Nicht jede Form von *Hate Speech* mündet in *Dangerous Speech*. So stellen die Anti-gay Proteste der Westboro Baptist Church auf militärischen Trauerfeiern (Tsesis 2013) *Hate Speech*, aber kein *Dangerous Speech* dar (Bruce-Lockhart 2013).

Das Modell des *Dangerous Speech* entstand im Kontext offener Konflikte in Afrika, die von Hassanstiftenden Reden zumindest mit-vorangetrieben wurden, wie z.B. der Ausbruch von Gewalt im Rahmen der kenianischen Wahlen 2007/2008. Die Gewalt zwischen den ethnischen Gruppen eskalierte damals im Zusammenhang mit massiven Wahlbetrugsvorwürfen. Die Unruhe hinterließ über tausend Todesopfer, dazu Szenen, die den Genozid in Ruanda in Erinnerung riefen (Gustafsson 2016; Gichuhi Kimotho und Nyaga 2016).

Anders als bei *Hateful Speech* handelt es sich im Fall von *Dangerous Speech* nicht unbedingt um eine affektuelle Handlung, sondern diese kann durchaus geplant und daher (zweck)rational erfolgen:

„Genocide and other forms of mass violence occur neither spontaneously nor abruptly. They follow a process of social conditioning to build up hatred and fear until those emotions become reflexive, and to place other human beings outside the ‚universe of moral obligation‘ (...).“ (Benesch 2014b, S. 6)

Studien zu *Dangerous Speech* versuchen dieses mit unterschiedlichen Forschungsdesigns und Untersuchungsmaterialien zu erfassen (Yanagizawa-Drott 2014; Straus 2007; Gichuhi Kimotho und Nyaga 2016; Gagliardone 2015). Das *Umati monitoring project* z.B. sammelte Beiträge in *Blogs*,

*Tweets, Online Newspaper Comments*, auf Facebook-Seiten sowie in anderen Onlinebeiträgen, um so den Medienwandel zu beobachten (Umati 2013).

Die Häufigkeit und die Übertragungsform dieser Art von *Hate Speech* variiert stark nach dem Kontext und dem Medium. Festgestellt wurde, dass 25 Prozent aller Beispiele von *Hateful Speech* offene Aufrufe zur Gewalt (töten, schlagen usw.) gegen Menschen aufgrund einer „Gruppen“-Zugehörigkeit enthielten. Fast alle SprecherInnen (94 Prozent) posteten zudem mit unter einem Namen (obwohl die ForscherInnen nicht exakt bestimmen konnten, ob es sich um echte Namen handelte). Die *Social Networking Sites* spielten hierbei eine wesentliche Rolle: 83 Prozent der hassgefüllten Aussagen wurden auf Facebook-Seiten gefunden und weniger als 5 Prozent bei Twitter (Benesch 2014b, 12f.).

Während in den Jahren 2007 und 2008 Botschaften hauptsächlich via SMS, E-Mail oder durch spezifische Websites verbreitet wurden, spielten soziale Netzwerke wie Twitter und Facebook 2013 die zentrale Rolle.

Auch wenn die Kommunikationssituation in offenen Konflikten anders aussieht, bietet dieser Ansatz große Vorteile gegenüber *Hateful Speech*-Ansätzen bzw. dem lokutionärem und dem illokutionärem Modell.

Zum einen ist es das Modell, das die Kommunikationssituation am ausführlichsten einbezieht. So werden SprecherInnen, Botschaft, Publikum und Medien berücksichtigt. Zum anderen ist es das erste *Hate Speech*-Modell, das Medien nicht nur einbezieht, sondern auch als Faktor, als eine möglicherweise determinierende Variable ansieht.

Außerdem deutet dieser Ansatz nur die Intention, sondern auch die Rationalität von *Hate Speech* (sowie des Genozids selbst) an. Er erkennt zuerst, dass Genozide eine Vorgeschichte haben und ausführlich geplant werden, was auf Zweckrationalität hindeutet. Last but not least zeigt das Modell, dass *Hate Speech* ausführlich begründet werden kann, um Gewalt gegen

eine Gruppe zu rechtfertigen und anzustiften (Leader Maynard und Benesch 2016).

Da dieser Ansatz sich aber alleinig auf *Hate Speech* konzentriert, das offene Gewalt hervorbringt, kann er nicht automatisch auf andere Formen von *Hate Speech* übertragen werden.

## 2.5 Hate Speech und benachbarte Begriffe

Das Übersehen der vorgestellten Wesensmerkmale von *Hate Speech* führt immer wieder dazu, dass *Hate Speech* mit wertbeladenen, abwertenden evaluativen Äußerungen, aber auch mit präskriptiven Aussagen mit negativen Konnotationen gleichgesetzt wird (Gagliardone et al. 2015, S. 7; Neumann und Arendt 2016). Bezieht sich eine negative Äußerung nicht (direkt oder indirekt) auf eine Kategorie von Menschen oder ist sie nicht intentional, dann handelt es sich nicht um *Hate Speech*. Steht im Kern der Botschaft keine Antinomie von Menschen auf Basis einer Kategorie, dann geht es auch nicht um diese Form kommunikativer Diskriminierung.

### 2.5.1 Beschimpfung, Beleidigung und Fauxpas

**Beschimpfungen** überschneiden sich mit *Hate Speech*, so dass man ebenfalls von *Hate Speech*-Beschimpfungen sprechen kann. Beschimpfungen haben mit *Hate Speech* die Tatsache gemeinsam, dass sie abwerten und ebenfalls auf Kategorisierung zurückgreifen **können**. *Hate Speech* unterscheidet sich von Beschimpfungen dadurch, dass es sich immer auf „Gruppen“ (kategorisierte Menschen) bezieht. Beschimpfungen unterscheiden sich von *Hate Speech*, da hierfür eine eigene Lexikalisierung unerlässlich ist (Carnaghi und Maass 2007). Darum entsprechen *Hate Speech*-Beschimpfungen oft der Anwendung von Ethnophaulismen oder *Racial Shurs*.

Ein anderer Unterschied liegt in dem Gefühlszustand der SprecherInnen. Beschimpfungen verlangen Erregungszustände (Stenzel 1986, S. 5), 5), so

dass ihre Bestimmungsgründe als affektiv/emotional einzustufen sind ((Weber 1988, S. 12). Beschimpfung ist also – in Anlehnung an Weber (1988) – nicht rational. *Hate Speech* verlangt wiederum nicht unbedingt einen Erregungszustand.

Beschimpfen hat zudem eine Ausdrucksfunktion. Daher kann Beschimpfung auf eine bloße Reihung von Schimpfwörtern reduziert werden (Stenzel 1986, S. 4). *Hate Speech* kann wiederum über diese Ebene hinausgehen und eine Darstellungsfunktion umfassen. Das sorgt dafür, dass Beschimpfungen expressive Sprechakte darstellen, während *Hate Speech* sich nicht auf einen Typ illokutionäre Akte beschränkt.

Auch wenn Intentionalität *Hate Speech* und Beschimpfungen einander nahe bringen, unterscheiden sich beide Formen hinsichtlich der Rationalität und der daraus folgenden Diskursqualität. Da Beschimpfungen als expressive Sprechakte fungieren, die eine Ausdrucksfunktion und damit den Zweck haben, Gefühlslagen auszudrücken, besitzen sie keine Diskursqualität (vgl. dazu Steenbergen et al. 2003). *Hate Speech* kann jedoch durchaus Begründungsrationalität besitzen.

**Beleidigungen** unterscheiden sich prinzipiell von *Hate Speech* dadurch, dass sie das individualisierte Pendant zu Diskriminierung darstellen. Während Diskriminierung eine Bewertung aufgrund einer Kategorisierung von Menschen darstellt, bildet Beleidigung die Bewertung auf Basis eines individuellen Merkmals der betreffende Person, das in keinem Zusammenhang mit einer sozialen Kategorie steht (Wagner 2001, 14; 119). Außerdem wird Beleidigung ausschließlich durch ihre Wirkung bestimmt. Es gilt das, was verletzt, und nicht das, was verletzend oder abwertend ist. Man kann sich aber verletzt fühlen, auch wenn der Inhalt nicht abwertend ist und vice versa (vgl. Unger 2013, S. 273).

Ein „Gruppen“-bezug tritt bei Beleidigungen höchstens dann auf, wenn sich ein Individuum wegen seiner vermeintlichen oder tatsächlichen Zuge-

hörigkeit zu einer „Gruppe“ abgewertet **fühlt**. Darin liegt auch der Unterschied zu Beschimpfungen: Eine Beleidigung liegt im perlokutionären Effekt bzw. auf der höheren Ebene und nicht an einem illokutionären Akt bzw. auf der Ebene der SprecherInnen, wie dies auch bei Beschimpfungen der Fall ist (vgl. Unger 2013).

**Fauxpas** können so wie *Hate Speech* durch rassistische, sexistische oder homophobe Äußerungen ausgelöst werden, sind aber a priori „a social blunder involving **unintentional** insult“ (Banerjee et al. 2011, S. 1887, Herv. L.S.).

Tabelle 3: Formen symbolischer Abwertungen

<b>Merkmale</b>	<b>Diskrimi- nierende Sprache</b>	<b>Hate Speech</b>	<b>Be- schimp- fung</b>	<b>Beleidi- gung</b>	<b>Fauxpas</b>
Öffentlichkeit	möglich	ja	möglich	möglich	ja
„Gruppen“-bezug/ Kategorisierung	ja	ja	möglich	möglich	möglich
Evaluation/ Wertung	ja	ja	ja	ja	ja
Intentionalität	möglich	ja	ja	möglich	nein
Begründungs- rationalität	möglich	möglich	nein	nein	nein
Lexikalisierung	möglich	möglich	ja	möglich	möglich
Performativität/ Wirkung	möglich	möglich	möglich	ja	ja

Quelle: eigene Darstellung

Ursache eines *Fauxpas* sind u.a. falscher Glaube und Unkenntnis. Solche „Fettnäpfchen“ lassen sich eher auf interkulturelle oder soziale Inkompetenz zurückführen. Darum ist Intentionalität, anders als bei *Hate Speech* kein Wesensmerkmal des *Fauxpas*. Haupteigenschaft von ***Fauxpas*** ist die Unangemessenheit (vgl. Molinsky 2005, S. 104).

Der Mangel an Intentionalität und der Ursache einer mangelnden interkulturellen oder sozialen Kompetenz hindern *Fauxpas* daran, eine höhere diskursive Qualität zu erreichen: „Fettnäpfchen“ lösen maximal Rechtfertigungen, aber keine Begründungen aus (vgl. dazu Kuhlmann 1999, S. 67). Das bedeutet, dass diese Äußerungen nicht im Vorfeld, sondern erst danach mit „Argumenten“ untermauert werden, wenn überhaupt.

### 2.5.2   Hate Speech und Hasspostings

Im Alltagsverständnis wird *Hate Speech* oft unter der Bezeichnung Hasspostings gefasst, d.h. einer Reihe von Formen symbolischer Gewalt, zu denen *Bullying*, *Stalking*, der Aufruf zu Straftaten oder auch Verleumdung gehören (Gagliardone 2015, S. 7; Neumann-Braun 2002).

Der erste Unterschied zwischen *Hate Speech* und Hasspostings bezieht sich auf den Inhalt, genauer gesagt, auf den „Gruppen“-bezug. Hasspostings beschränken sich nicht auf Kategoriebezogene Aussagen und zumeist ist damit lediglich *Cyber Hate* gemeint (Wettstein 2016; 2016).

Nicht alle Formen von *Cyber Hate* stellen jedoch *Hate Speech* dar. *Cyber Bullying* z.B. besitzt keine Schnittmenge mit dem letzteren. *Cyber Bullying* ist aggressiv, intentional, hat aber keinen Kategoriebezug sondern richtet sich direkt auf ein Individuum und stellt eine repetitive Handlung dar (Saleem et al. 2016).

Hasspostings beschränken sich zudem auf eine Mediengattung (Online-Medien), während *Hate Speech* nicht einmal medienvermittelt sein muss.

Auch in den Fällen Kategorienbezogener Hasspostings stellen diese in der Regel eine spezielle Form von *Hate Speech* dar – *Hateful Speech*. Es ist also nicht verwunderlich, dass die Erforschung von Hasspostings oft in der Forschungstradition des Deliberationsmodells steht (Neumann und Arendt 2016; Wirth und Wettstein 2013; Rowe 2014; Coe et al. 2014), in der *Hate Speech* auf ein emotionales Phänomen reduziert wird, das mit Rationalität (nach dem diskursivem Modell von Habermas 1981a) überwunden werden soll. Setzt man *Hate Speech* mit Hasspostings in diesem Kontext gleich, dann ist es allein ein Problem der Diskursqualität. Folgt man einem solchen Modell, stellt die bewusste kommunikative Herstellung von Minderwertigkeit von Menschen kein *Hate Speech* dar und der Blick beschränkt sich nahezu ausschließlich auf *Hateful Speech*.

Wie bereits analysiert, wird diese Form von *Hate Speech* hauptsächlich durch ungebildete und/oder unerfahrene öffentlichen SprecherInnen geäußert. Es ist äußerst unwahrscheinlich, dass hochrangige PolitikerInnen Schimpfwörter gegen Minderheiten verwenden oder offen zur Gewalt gegen sie aufrufen.

Die Fokussierung der *Hate Speech*-Debatte auf Hasspostings hat als Nebenwirkung, dass das Problem hauptsächlich thematisiert wird, wenn die SprecherInnen keine *Public Figures* sind und daher in einer ungünstigen Machtposition stehen. Die Chancen Nicht-öffentlicher SprecherInnen eine Wertgeneralisierung zu bewirken, ist jedoch viel begrenzter als die erfahrenen Öffentlichkeitsakteure. Daher haben sie im Vergleich zu PolitikerInnen z.B. geringere Chancen, solche Ansichten zu legitimieren und ein Konsens mit ihren diskriminierenden Beschimpfungen oder Drohungen zu bewirken.

*Hateful Speech*, wie im Fall von Hasspostings, hat andere Ursachen als bewusstes, (zweck-)rationales *Hate Speech*. Nicht jeder Hassposter ist ein Überzeugungstäter. Auch die Ursachen für ihre Handlungen können viel-

fältig sein. Erjavec und Kovačič (2012) interviewten AutorInnen von Hasspostings auf Nachrichtenseiten in Slowenien und identifizierten unterschiedliche Typen von *Hateful speakers*:

„Different types of commentators receive different benefits from engaging in hate speech commentary, although they all get a feeling of recognition and empowerment; as all of them said, in this way, their voice is “heard” and it “counts for something.” Soldiers and believers develop a feeling of participating in a direct political fight, watchdogs get a feeling of social engagement and subversion of the existent social situation and players get thrill and fun.“ (Erjavec und Kovačič 2012)

Der modus operandi der *Social Networking Sites* ermöglicht es nicht einmal zu sagen, dass Hasspostings tatsächlich menschliche Handlungen darstellen, denn sie können auch automatisch etwa durch sogenannte *Social Bots* generiert werden (Ferrara et al. 2016; Mitter et al. 2013)

Das soll nicht heißen, dass Hasspostings keine Relevanz für *Hate Speech* hätten. Hasspostings erzeugen kein Prestige, aber durchaus Einfluss, denn sie generieren Konflikte bzw. Kontroversen. Obwohl Hasspostings nur einen kleinen Anteil ausmachen und meistens von Nicht-öffentlichen SprecherInnen geäußert werden, lösen sie online die meisten Reaktionen in den sozialen Netzwerken aus (Gagliardone et al. 2015).



### 2.5.3 Hate Speech und Political Incivility

Der Anteil von Hasspostings, die keinen „Gruppen“-bezug aufweisen, stellt ein anderes, ebenfalls kommunikations- und medienwissenschaftlich relevantes Phänomen dar. Beschimpfungen von PolitikerInnen und/oder zwischen NutzerInnen ist eine Frage der *Political Incivility*, d.h. des Verhaltens der DiskutantInnen/politischen AntagonistInnen, welche einen freien und respektvollen Austausch von Ideen in politischen Deliberationsprozesse konterkariert (Stryker et al. 2016). Dazu gehören zum Beispiel abwertende Spitznamen, die Verwendung eines vulgären Wortschatzes sowie intentionale Behauptungen von Unwahrheiten.

*Political Incivility* beeinträchtigt die Diskursqualität politischer Debatten. Die Verwandlung politischer Deliberation in einen Schlagabtausch birgt Gefahren für moderne Demokratien, da sie deren Legitimationsquelle – die Deliberation und die Bürgerpartizipation – selbst in Frage stellt. Darüber hinaus offenbaren derartige Beschimpfungen, Gewaltandrohungen und -rechtfertigungen auch eine politische Polarisierung. Mangelnder Respekt und prekäre Argumentationen schüchtern zudem NutzerInnen ein und verhindern dadurch, dass sich diese äußern (Stryker et al. 2016; Russmann 2015).

Auch wenn beide Phänomene hohe soziale Relevanz besitzen ist *Political Incivility* nur in Ausnahmefällen mit *Hate Speech* gleichzusetzen. *Political Incivility* ist eine Frage der politischen Antagonismen. Sie resultiert daraus, dass zwei oder mehrere Akteure sich gegeneinander positionieren. Hate Speech ist dagegen eine Frage von Antinomien: Zwei Kategorien von Menschen werden als Gegensätze definiert. Es handelt sich also um keine Frage von Positionierung. Homophobisches *Hate Speech* resultiert z.B. nicht daraus, dass sich Homosexuelle oder Heterosexuelle gegeneinander positionieren, sondern daraus, dass sie als Gegensätze und Feinde definiert werden.

Tabelle 4: Formen von *Political Incivility* (Auszug)

<b>Variable</b>	<b>Definition</b>
Vulgarität	Verwendung obszöner oder vulgärer Sprache im politischen Diskurs
Verteufelung	Verwendung von mystisch-religiösen Metaphern
Group derogatory labels (z.B. Ethnophaulismus)	Verwendung von sexistischer, ethnischer oder religiöser Sprache in der politischen Debatte
Eigene Lexikalisierung	Anwendung abwertender Ausdrücke, um Dissens oder Verachtung für die politischen GegnerInnen auszudrücken
Drohung	Androhung materieller oder physischen Schadens an die politischen GegnerInnen
Anstiftung	Ermutigung anderer in der politischen Diskussion, den politischen GegnerInnen materiellen oder physischen Schaden zuzufügen
Irreführung	Die intentionale Verwendung falscher Aussagen (Lügen) oder die Behauptung irreführender Angaben
Irreführung	Unterbindung der Teilnahme von politischen GegnerInnen an der Debatte
Zensur	Anwendung sprachlicher Mittel, um Sachverhalte unverhältnismäßig zu übertreiben
Übertreibung	Angriffe auf persönlichen Eigenschaften der politischen GegnerInnen
Übertreibung	Anwendung sprachlicher Mitteln, um Sachverhalte unverhältnismäßig zu übertreiben

Quelle: eigene Darstellung in Anlehnung an Stryker et al. 2016

Es kann zum Antagonismus führen, aber dieser besteht aus Homophoben und Nicht-Homophoben. Bei *Political Incivility* sind das Ziel des Angriffs die SprecherInnen der anderen Seite. *Hate Speech* dagegen bedeutet nicht unbedingt mitreden, sondern auch darüber reden, d.h. die Angegriffenen sind nicht unbedingt diejenigen auf der anderen Seite.

Anders als *Political Incivility* hat *Hate Speech* immer einen „Gruppen“-bezug. Wenn z.B. der US-amerikanische Prediger Pat Robertson in einer Fernsehsendung seinen Wunsch äußert, dass die amerikanische Geheimdienste den damaligen Präsident Venezuelas Hugo Chávez ermorden, dann handelt es sich um Aufstachelung zur Gewalt, aber nicht um *Hate Speech*<sup>5</sup>. Stachelt man aber zum Mord an PolitikerInnen aufgrund von deren Zugehörigkeit zu einer „Gruppe“ auf, dann handelt es sich nicht nur um einen Aufruf zur Gewalt, sondern auch um *Hate Speech*.

So geschehen im Fall eines österreichischen Kommunalpolitikers. Auf seiner Facebook-Seite veröffentlichte der damalige FPÖ-Politiker ein Foto einer türkischstämmigen Abgeordneten neben einer mutmaßlich gesteinigten Frau (APA 2015). Zugleich beschimpfte er die Politikerin unter anderem als „wahnsinnige Dreckschleuder aus der Türkei“. In diesem Fall handelt es sich um *Political Incivility* und *Hate Speech*.

Wie hoch ist aber der Anteil von Hasskommentaren, die sich auf *Political Incivility* zurückführen lassen? Wirth und Wettstein (2013) untersuchten die Kommentare der wichtigsten Nachrichtenseiten in der Schweiz und stellten einen Anteil von 8,3 Prozent an persönlichen Angriffen fest. Coe, Kenski et al. (2014) ermittelten 22 Prozent von *Politically Incivility*-Kommentaren auf den Webseiten der Tageszeitungen aus Arizona. Die häufigste Form war *Derogatory Name-Callings*. Allerdings stellten die ForscherInnen keinen Zusammenhang zwischen *Political Incivility* und Irrationalität her, da ein großer Anteil der aggressiver Kommentare ebenfalls

---

<sup>5</sup> <http://edition.cnn.com/2005/US/08/23/robertson.chavez/>

empirische Evidenzen enthielten (Coe et al. 2014, S. 674). Rowe (2014) fand über 30 Prozent derartiger Kommentare auf der Webseite und auf dem Facebook-Account der *Washington Post*.

Allerdings stellt sich die Frage, wann ist eine Botschaft *politically incivil*. Darauf gibt es keine einheitliche Antwort. Stryker et al. (2016) befragten US-amerikanischen StudentInnen nach ihrer Wahrnehmung von *Political Incivility* anhand unterschiedlicher *Civility Indicators*. Während über 80 Prozent *Derogatory Group Labels* extrem unzivilisiert fanden, stuften nur 53 Prozent Beleidigungen als extrem unzivilisiert ein (Stryker et al. 2016).

#### 2.5.4 Hate Speech und Rechtspopulismus

Haupteigenschaften des politischen Phänomens Populismus ist eine Komplexitätsreduktion, die auf dem Konzept des Volkes als einer Einheit, das den Gegenpol zu der Elite bildet (Heinisch 2013, S. 49). Der Antagonismus zwischen Volk und Elite bildet der Kern des Populismus (vgl. Wettstein et al. 2016, S. 15).

Die politische Plattform von Rechtspopulisten profitiert und bedient sich gleichzeitig der Ängste des Publikums. Rechtspopulisten verwenden/verbreiten nicht nur diskriminierende Inhalte. Sie müssen das auch in einer Kommunikationsform tun, die es ihnen ermöglicht, die Ängste zu schüren. Das erfolgt mitunter durch eine *Scapegoating*-Rhetorik (Wodak 2015, S. 1), in der eine „Gruppe“ (z.B. MuslimInnen, MigrantInnen oder Flüchtlinge) für alle gesellschaftliche Probleme verantwortlich gemacht wird, egal, ob es sich um Arbeits-, Wohn- oder Schulpolitik handelt.

Dadurch kanalisieren Rechtspopulisten Meinungen und versuchen WählerInnen anzusprechen. Mehr als das: Durch die bewusste und geplante Anwendung Gruppenbezogener Inhalte versprechen sie sich, WählerInnen zu mobilisieren, die Medienagenda mit ihren Themen zu besetzen und damit zugleich ihre politische GegnerInnen dazu zu bringen, sich gegenüber

den von ihnen gesetzten Themen zu äußern (Sponholz 2016a). *Hate Speech* stellt also ein zentrales Element der politischen Kommunikation von Rechtspopulisten dar.

Allerdings ist *Hate Speech* nicht für den Populismus allgemein, sondern nur im Fall der VertreterInnen des rechten politischen Spektrums, seien es Populisten oder Extremisten, von Relevanz. Dies ist darin begründet, dass Rechtspopulisten den Antagonismus zwischen Elite und Volk auf einer identitären Politik aufbauen. Während sich Linkspopulismus durch Internationalismus und Post-Nationalismus kennzeichnet, ist der Rechtspopulismus nationalistisch/chauvinistisch und beruht auf der Idee von Volk als nativistischem Körper (Wodak 2015, S. 8).

Die Gleichsetzung von *Hate Speech* und Populismus erfolgt vor allem in mittel- und osteuropäischen Ländern aufgrund der politischen Spielarten, die sich in diesem Teil des Kontinents entwickelt haben. Anders als in Südeuropa und in Lateinamerika ist die rechte Ausprägung in diesen Ländern so stark, dass sogar in der Fachliteratur auf die Bezeichnung der ideologischen Lager verzichtet wird (Decker 2006).

So wird Populismus allgemein zum Synonym für VertreterInnen des rechten politischen Spektrums. Diese Reduzierung führt dazu, dass der Populismus selbst und nicht die Rahmen der rechten oder der linken Ausprägung als Ideologie bezeichnet wird (Rensmann 2006; Wettstein et al. 2016)

Für Laclau (2005) ist Populismus dagegen nicht eine Ideologie, sondern eine „mode of articulation“ sozialer, politischer oder ideologischer Inhalte. Der populistische Artikulationsrahmen hat dadurch eine Orientierungsfunktion und verwandelt sich in eine Vermittlungs- bzw. Kommunikationsstrategie. In dieser Hinsicht zeigen Flecker und Kirschenhofer (2007, S. 30), dass Sichtweisen und Erklärungen, die auf ethnischer Differenzierung, Ausgrenzung von Randgruppen oder Spaltung abstellen, einfache

Lösungen für komplexe Probleme bieten. Auch Priester (2012) verdeutlicht, dass Populismus selbst keine Ideologie oder Weltanschauung ist. Die kommunikative Strategie von Populisten kann sogar für elitäre politische Programme verwendet werden:

„Populismus betreibt keine bloße Aufwertung des Volkes, sondern eine Umpolung der Wertigkeiten von Volk und Elite und ist nur in einem instrumentellen Sinne anti-elitär“ (Priester 2012, S. 5).

Anti-Elitismus ist ein Wesensmerkmal des Populismus als „mode-of articulation“ oder instrumentelle Kommunikation. Für *Hate Speech* ist es das dagegen nicht. Im Gegenteil: Diese Kommunikationsform von Diskriminierung kann durchaus elitär sein, wie der Fall Sarrazin illustriert. So war eine seiner Hauptforderungen aus Berlin, der „Hauptstadt der Transferleistungen“ eine „Metropole der Eliten“ zu machen (Berberich 2009; Friedrich 2011). Im Kern von *Hate Speech* steht eine Antinomie zwischen Menschen auf Basis einer Kategorie (Geschlecht, Herkunft, Phänotyp, Religion, sexuelle Orientierung usw.), im Rechtspopulismus dagegen eine Antinomie zwischen Elite und Volk.

*Hate Speech* kann zudem in unterschiedlichen Systemen vorkommen und beschränkt sich nicht auf die Politik. Religiöse FührerInnen und Gruppierungen können sich dessen bedienen, wie der Fall von Westboro Baptist Church-Proteste in den USA zeigt. Auch im Mediensystem hat *Hate speech* einen festen Platz, ist es doch für einen großen Teil des Umsatzes im Verlagswesen verantwortlich, wie wie die hier untersuchten Fälle von Fallaci und Sarrazin illustrieren.

Hate Speech in den Massenmedien

Theoretische Grundlagen und empirische Umsetzung

Sponholz, L.

2018, XX, 454 S. 22 Abb., Softcover

ISBN: 978-3-658-15076-1