

# Sylvia Jaki<sup>1</sup>

<sup>1</sup>Universität Hildesheim

[jakisy@uni-hildesheim.de](mailto:jakisy@uni-hildesheim.de)

## **Charakteristika von Hate Speech in sozialen Medien und Herausforderungen für deren automatische Erkennung**

Galt Hate Speech vor wenigen Jahren noch als neuer, faktisch unbeforschter Gegenstand der Linguistik und anderer Disziplinen, liegen heute einerseits zahlreiche Untersuchungen zu verschiedenen Ausprägungen von Hate Speech in sozialen Medien vor (vgl. für einen Forschungsüberblick Jaki in Vorb.). Andererseits bestehen intensive computerlinguistische Bemühungen zur automatisierten Erkennung von Hate Speech, zum Beispiel in Form sogenannter *Shared Tasks* wie *GermEval* (Struß et al. 2019) oder *HASOC* (Modha et al. 2021).

In diesem Vortrag sollen die sprachlichen Charakteristika von Hate Speech überblickshaft erläutert und zu deren Exemplifizierung Beispiele aus vergangenen Fallstudien (Jaki & De Smedt 2019, Jaki et al. 2019) sowie den französischsprachigen Daten des *European Observatory of Online Hate* (eooH.eu) herangezogen werden. Hierbei wird schwerpunktmäßig auf die Bereiche Xenophobie und Misogynie bzw. Antifeminismus eingegangen. Wengleich diese (insbesondere lexikalischen) Charakteristika auch für die automatisierte Erkennung von Hate Speech eine wichtige Rolle spielen, wird gezeigt, warum hierdurch die Palette an Phänomenen, die mit Hate Speech einhergeht, bei weitem nicht abgedeckt ist.

Trotz des Fokus auf sprachliche Charakteristika möchte der Beitrag auch dafür plädieren, Hate Speech als multimodale Kommunikationsform aufzufassen, in der vor allem bildliche Elemente, zum Beispiel in Form von Memes, eine zunehmend wichtige Rolle spielen. Lange Zeit konzentrierte sich sowohl die deskriptive Forschung als auch die automatische Erkennung fast ausschließlich auf sprachliche Charakteristika, während Text-Bild-Beziehungen weitestgehend außer Acht gelassen wurden. Der mittlerweile erfolgte *multimodal turn* in deskriptiven Arbeiten hat auch die computerlinguistische Forschung zum Thema bereits erreicht (Kiela et al. 2020).

**Sektion 7 : Le discours de haine en français : manifestations verbales de la haine dans les médias sociaux**

## Literaturverzeichnis

- Jaki, Sylvia. in. Vorb. Hate Speech in sozialen Medien. Ein Forschungsüberblick aus Sicht der Sprachwissenschaft. In Sylvia Jaki & Stefan Steiger (eds.), *Digitale Hate Speech - Interdisziplinäre Perspektiven auf Erkennung, Beschreibung und Regulation*. Wiesbaden: Springer.
- Jaki, Sylvia & Tom De Smedt. 2019. *Right-wing German hate speech on Twitter: Analysis and automatic detection*. [arxiv.org/abs/1910.07518](https://arxiv.org/abs/1910.07518).
- Jaki, Sylvia, Tom De Smedt, Maja Gwózdź, Rudresh Panchal, Alexander Rossa & Guy De Pauw. 2019. Online hatred of women in the Incels.me forum: Linguistic analysis and automatic detection. *Journal of Language Aggression and Conflict* 7(2). 240–268. [doi.org/10.1075/ilac.00026.jak](https://doi.org/10.1075/ilac.00026.jak).
- Kiela, Douwe, Hamed Firooz, Aravind Mohan, (...) & Davide Testuggine. 2020. The hateful memes challenge: Detecting hate speech in multimodal memes. *Advances in Neural Information Processing Systems* 33. 2611–2624.
- Modha, Sandip, Thomas Mandl, Gautam Kishore Shahi, (...) & Marcos Zampieri. 2021. Overview of the HASOC Subtrack at FIRE 2021: Hate Speech and Offensive Content Identification in English and Indo-Aryan Languages and Conversational Hate Speech. In *FIRE 2021: Forum for Information Retrieval Evaluation*, ACM. <https://dlnext.acm.org/doi/abs/10.1145/3503162.3503176>
- Struß, Julia Maria, Melanie Siegel, Josef Ruppenhofer, Michael Wiegand & Manfred Klenner. 2019. Overview of GermEval Task 2, 2019 shared task on the identification of offensive language. In *Proceedings of the 15<sup>th</sup> Conference on Natural Language Processing (KONVENS) Nürnberg/Erlangen*. <https://doi.org/10.5167/uzh-178687>