

# Fanfiction Semantics – Eine quantitative Analyse sensibler Themen in deutscher Fanfiction

**Häußler, Julian**

julian.haeussler@stud.tu-darmstadt.de  
Technische Universität Darmstadt, Deutschland

## Hintergrund

Als Fanfiction können all jene Texte bezeichnet werden, die zur Veröffentlichung in eigens dafür eingerichteten Internetforen bestimmt sind und in „appropriativ-derivativer bzw. [...] transformativer“ (Stemberger 2021, 10) Weise Bezug auf Mainstreammedien (meist Romane, Serien oder Filme) nehmen. Für die Autor\*innen von Fanfiction, die meist unter Pseudonym schreiben, ist es somit möglich, in einem geschützten Raum (der Fancommunity) neben ersten Schreibversuchen, zu ihren Lieblingsfiguren auch das Verfassen von die Bezugsmedien kontrastierenden Erzählungen zu wagen. Catherine Tosenberger beschreibt letztere Kategorie als “a freedom especially felt with regard to non-normative and taboo forms and representations of sexuality” (Tosenberger 2014, 17). Fanfiction, die Literatur als Bezugsquelle gewählt haben, bauen dabei meistens auf großen Mainstream-Jugendbuchreihen aus dem angelsächsischen Raum auf. So sammeln sich in der deutschsprachigen Community auf fanfiction.de unter der Kategorie ‚Bücher‘ die meisten Autor\*innen in den Fandoms zu Harry Potter (über 55.000 Texte gesamt), Bis(s) (knapp 14.000) und Herr der Ringe (als allg. Gruppe Mittelerde definiert mit über 8.000 Texten). Praktisch befassen sich die Autor\*innen oft mit Rekombinationen des Figurenarsenals der Originaltexte. Eine wichtige Rolle spielt dabei das Stereotyp der Slashes (männliche Figuren, die in eine romantisch-sexuelle Beziehung gesetzt werden; vgl. Brottrager et al. 2022). Die Moralvorstellungen der Originale können dabei auch übergangen werden, wenn zum Beispiel der Schüler Harry mit seinem Lehrer Severus Snape eine (romantisch-sexuelle) Beziehung eingeht.

## Ziele des Projekts

Das Masterprojekt „Fanfiction Semantics – A Quantitative Analysis of Sensitive Topics in German Fanfiction“ hat zum Ziel, diese Kontrastierungen in verschiedenen Fandoms zu betrachten. Gefragt wird dabei einerseits, wie virulent sensible Themen (sensitive topics) in Fanfiction sind und andererseits, wie diese Themen Wortbedeutungen beeinflussen. Als sensibel werden dabei

jene Themen definiert, die in Originaltexten mehrheitlich ausgespart und von Fanfiction-Autor\*innen demonstrativ eingeführt werden (v. a. extreme Gewalt und sexuelle Inhalte). Das Korpus entstammt einem seit 2020 andauernden Webscrapings am LitLab der Technischen Universität Darmstadt (vgl. Weitin 2022). Hier wurden alle in diesem Zeitraum bearbeiteten Texte sowie die von der Plattform definierten Metadaten (neben Selbstangaben der Autor\*innen auch Angaben zu Altersbeschränkung der Texte und der eigenen Definition von Genre) heruntergeladen. Das Projekt bearbeitet darauf aufbauend in zwei Schritten zunächst die Metadaten zu den größten Fandoms im Bereich ‚Bücher‘ und wertet anschließend die dazugehörigen Texte quantitativ aus. Der untersuchte Ausschnitt des Fanfiction-Korpus umfasst über 8.000 Einzelltexte mit knapp 130 Mio. Tokens.

## Methodik

1) Im ersten Schritt, bei der Analyse der Metadaten, kann beispielsweise eine Verteilung davon, welche der Kategorien zur Altersbeschränkung für wie viele Texte verwendet wurde, einen Hinweis auf die Unterschiede zum Bezugsmedium liefern. Die auf fanfiction.de gesetzte Kategorisierung reicht dabei von Texten, die ab 6 Jahren geeignet sein sollen, bis hin zu solchen, die als „entwicklungsbeeinträchtigend“ (fanfiction.de) gelten. Entsprechende Verteilungen werden auch zwischen Fandoms verglichen.

2) Die Analyse der Texte erfolgt im folgenden Schritt mithilfe von Word Embedding Modellen (hier word2vec; vgl. Mikolov et al. 2013). In einem Word Embedding Modell können Ähnlichkeiten zwischen verschiedenen Schlüsselwörtern effizient verglichen werden. Die algorithmische Grundlage bietet dabei die Verwendung der Wörter in einem festgesetzten Kontextfenster (vgl. Distributionshypothese nach Firth; vgl. Firth 1957). Durch den Abgleich von Schlüsselwörtern, die auf verschiedene Themen verweisen, können Schnittmengen und Unterschiede untersucht werden. Die Schlüsselwörter wurden hierzu aus den Wortfeldern Gewalt und sexuelle Inhalte gewählt, welche Wörter unter den (hier) 30 ähnlichsten zu einem Zielwort zu finden sind, beschreibt in gewisser Weise wie dieses Zielwort verwendet wird (vgl. Abb. 1). Für das Subkorpus der *Harry Potter*-Reihe werden zudem dieselben Operationen auch für die Originaltexte durchgeführt. Abschließend wird ein Verfahren der Word Embedding basierten Sentiment Analyse nach SentiArt (vgl. Jacobs 2019; Brottrager et al. 2022) durchgeführt, welche anhand der Variablen Valence (emotionale Wertigkeit) und Arousal (emotionales Erregungspotential) Schlüsselwörter beschreibt.

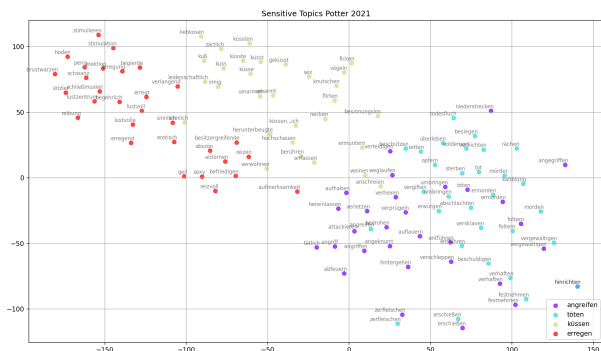


Abb. 1: t-SNE-Visualisierung der ähnlichsten Wörter zu vier Schlüsselwörtern in einem Word2Vec-Modell zu *Harry Potter*-Fanfiction.

tion“. *Children’s Literature Association Quarterly* 39 (1): 4–27. <https://doi.org/10.1353/chq.2014.0010>.

**Weitin, Thomas.** 2022. „Litlab“. <https://www.linglit.tu-darmstadt.de/institutlinglit/mitarbeitende/weitin/litlab/index.de.jsp> (zugegriffen: 03. August 2022).

## Ausblick

Das Poster soll die Ergebnisse der oben beschriebenen Analysen grafisch darstellen und einen Eindruck über die thematische Zusammensetzung der Fanfiction-Korpora bieten. Das zugrundeliegende Korpus kann aus urheberrechtlichen Gründen nicht in der verwendeten Form veröffentlicht werden. Dennoch ist es beabsichtigt, nach Abschluss des Projekts alle verwendeten Codes auf GitHub bereitzustellen, um Transparenz zu schaffen und die Methodik für andere Projekte nachnutzbar zu machen (orientiert an der in den FAIR-Prinzipien definierten Wiederverwendbarkeit (vgl. GO FAIR 2022)). Anonymisierte Metadaten tabellen sowie die Word Embedding-Modelle sollen zudem auch zugänglich gemacht werden.

## Bibliographie

**Brottrager, Judith, Joël Doat, Julian Häußler, und Thomas Weitin.** 2022. „Character Shifts in Harry Potter Fanfiction“. Herausgegeben von Thomas Weitin. *LitLab Pamphlet*, Nr. #10.

**fanfiktion.de.** 2022. „Richtlinien für die Alterskennzeichnung“. <https://www.fanfiktion.de/p/ageadvice/0> (zugegriffen: 03. August 2022)

**Firth, John.** 1957. „A Synopsis of Linguistic Theory, 1930–1955“. In *Selected Papers of J.R. Firth 1952–1959*, herausgegeben von Frank Palmer, 168–205. London: Longman.

**GO FAIR.** 2022. „FAIR Principles“. <https://www.go-fair.org/fair-principles/> (zugegriffen: 03. August 2022).

**Jacobs, Arthur M.** 2019. „Sentiment Analysis for Words and Fiction Characters From the Perspective of Computational (Neuro-)Poetics“. *Frontiers in Robotics and AI* 6. <https://doi.org/10.3389/frobt.2019.00053>.

**Mikolov, Tomas, Kai Chen, Greg Corrado, und Jeffrey Dean.** 2013. „Efficient Estimation of Word Representations in Vector Space“. *arXiv*. <https://arxiv.org/abs/1301.3781>.

**Stemberger, Martina.** 2021. *Homer meets Harry Potter: Fanfiction zwischen Klassik und Populärkultur*. Dialoge. Tübingen: Narr Francke Attempto.

**Tosenberger, Catherine.** 2014. „Mature Poets Steal: Children’s Literature and the Unpublishability of Fanfic-