

# Dunkelgrün, blassgrün, fenchelgrün oder: Über die Konkretisierung des Vokabulars im deutschsprachigen Roman (1760–1920)

**Hilger, Agnes**

agnes.helene.hilger@gmail.com

Julius-Maximilians-Universität Würzburg, Deutschland

## Hintergrund

Im Jahr 2012 entdecken Ryan Heuser und Long Le-Khac, dass eine Reihe von semantisch verwandten Wörtern in englischsprachigen Romanen über das 19. Jahrhundert hinweg immer häufiger verwendet wird (vgl. Heuser/Le-Khac 2012). Diese Wörter sind tendenziell konkret.<sup>1</sup> Sie bezeichnen Körperteile wie *finger* oder *hair* oder Farben wie *red* oder *scarlet*. Ted Underwood zeigt in einer späteren Untersuchung, dass dieser Trend bereits 1760 einsetzt. Er weist angesichts des mehrere Epochen umfassenden Anstiegs auf eine Lücke im bisherigen literaturgeschichtlichen Wissen hin.

Eine Masterarbeit, die dem Poster zugrunde liegt, verfolgte das Ziel, die Beobachtungen von Heuser, Le-Khac und Underwood zunächst versuchsweise für die deutschsprachige Literatur nachzuvollziehen und so dann eine erste Eingrenzung derjenigen Bereiche zu leisten, die von der Entwicklung betroffen sind. Die Ergebnisse sollen hier vorgestellt werden.

## Korpus und Methode

Das Korpus basiert auf den bei TextGrid und Projekt Gutenberg digital zur Verfügung gestellten Texten (vgl. Neuroth u.a. 2015; Reuters o.J.). Es enthält 1147 zwischen 1760 und 1920 erschienene Romane. Diese sind jedoch nicht gleichmäßig über den Zeitraum verteilt (s. Figure 1). Die Unausgewogenheit soll in anschließenden Arbeiten angegangen werden.

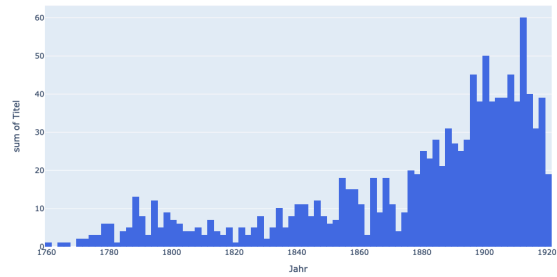


Figure 1: Übersicht über die im Korpus enthaltenen Romane pro Jahr.

Um die Wortfrequenzen zu ermitteln, wurden zunächst Wortlisten erstellt. Als Heuristik diente eine sehr grobe Einteilung in drei Gruppen: 1) Informationen zu Figuren, 2) Informationen zu Räumen und 3) Informationen zu Beschaffenheit und Material. Auf Basis dieser Unterscheidung wurde eine Liste von 31 Wortfeldern erstellt.<sup>2</sup> Anschließend wurden auf Basis von GermaNet Listen von zu diesen Wortfeldern gehörenden Wörtern erstellt (vgl. Henrich/Hinrichs 2010). Ein Beispiel für ein solches Wortfeld sind Farbwörter. Ausgehend vom Knoten *farbspezifisch* wurden dessen Hyponyme – zum Beispiel *grün* – und wiederum dessen Hyponyme – beispielsweise *dunkelgrün* und *blassgrün* – extrahiert.

Um der historischen Sprachstufe und der Domäne Roman gerecht zu werden, wurden die Wortlisten anschließend mit einem Word-Embedding-Modell erweitert. Dafür wurde ein auf CommonCrawl trainiertes FastText-Modell auf dem Roman-Korpus weitertrainiert (vgl. Bojanowski 2016). Aufgrund guter Performance in ähnlichen Tasks schien ein solches FastText-Modell ausreichend (vgl. Ehrmanntraut u.a. 2021). Um die Wortlisten zu erweitern, wurden zu den extrahierten Wörtern abhängig von der Länge der Liste die zwei bis zwanzig nächsten Nachbarn ermittelt und, sofern nicht schon vorhanden, der Liste angehängt. Neben Wörtern wie *grün* oder *dunkelgrün* enthielt die Liste nun auch sehr spezifische wie *fenchelgrün*. Anschließend wurden die Listen von Hand bereinigt.

Die Romane wurden mit dem Python-Paket spaCy lemmatisiert und für jedes Wortfeld die zugehörigen Wortfrequenzen berechnet (vgl. Honnibal/Montani 2017).

## Ergebnisse

Nimmt man die Wortfrequenzen aller 31 konkreten Wortfelder zusammen, ergibt sich ein signifikanter Anstieg (Mann-Kendall-Test,  $\alpha=0,01$ ,  $p=2,22e-16$ ). In Figure 2 repräsentiert jeder Punkt einen Roman, die y-Achse gibt jeweils die Wortfrequenz an. Im Korpus gibt es also einen ähnlichen Trend wie in den englischsprachigen Texten.

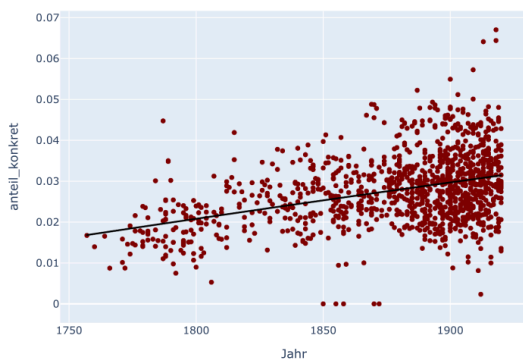


Figure 2: Frequenzen der konkreten Wörter pro Roman und Jahr.

Insbesondere bei den Wortfeldern, die Figuren, Gebäude und Innenräume beschreiben, gibt es signifikante Anstiege. Bei den Wortfeldern, die der Darstellung von Naturräumen (z.B. die Wortfelder *Gartenanlage*, *Gewächs*, *Bewaldung*) dienen dürften, gibt es dagegen keine signifikanten Trends. Figure 3 gibt einen Überblick über die Ergebnisse zu den untersuchten Wortfeldern ( $\alpha=0,01$ ).

Figure 3: Übersicht über die Trends für alle untersuchten Wortfelder.

	Anstieg	Rückgang	Kein Trend
1) Räume	Bau/Gebäude Gebäudeteil Zimmer Einrichtungsgegenstand/ Möbel Heimtextilie Haushaltsgegenstand/ Haushaltsprodukt Dekorationsgegenstand/ Ziergegenstand Gras/Grünfläche	Pflanzenteil	Gartenanlage/ Grünanlage Weg Gewächs/Pflanze Bewaldung/Wald Kunstobjekt Bild
2) Figuren	Körperteil Bekleidung/Kleidung Bekleidungsteil/ Kleidungsteil Aufmachung/Outfit Aussehensspezifisch Tasche		Geschmeide/Schmuck
3) Beschaffenheit	Gewebe/Stoff/Textil Farbspezifisch Helligkeitsspezifisch Oberflächenspezifisch Muster/Musterung Formspezifisch Geruch		Ornament/Verzierung Holz

## Fazit

Die Ergebnisse der Arbeit deuten auf eine grundlegende Veränderung im untersuchten Korpus hin, die von der germanistischen Literaturgeschichte bislang nicht wahrgenommen wurde: Die Art und Weise, in der Romane ihre fiktive Welt physisch gestalten, wandelt sich über einen mehrere Epochen umfassenden Zeitraum hinweg erheblich. Zudem ermöglichen die Ergebnisse der Arbeit eine erste Differenzierung. Besonders betroffen scheinen Informationen über das physische Erscheinungsbild von Figuren und Orten, an denen Figuren le-

ben. Eine anschließende Arbeit soll diese Ergebnisse konkretisieren.

## Fußnoten

1. Die Zuschreibung ‚konkret‘ ist als Heuristik zu verstehen. Ausschlaggebend ist für die Zuordnung anders als in der sprachwissenschaftlichen Abgrenzung von Konkreta und Abstrakta nicht die Gegenständlichkeit, sondern die Möglichkeit der sinnlichen Wahrnehmbarkeit. Dabei orientiere ich mich an der Definition, die Sabine Schulte im Walde und Maximilian Köper für ihren Datensatz benutzen (vgl. Köper/Schulte im Walde 2016).
2. Der Begriff ‚Wortfeld‘ ist hier weit gefasst und bezeichnet anders als in der Sprachwissenschaft keine Gruppe synonymen Wörter, sondern eine Gruppe von semantisch verwandten Wörtern, die sich unter einen Oberbegriff subsumieren lassen. Für die Wortfelder, die Gegenstände umfassen, fällt der Begriff daher mit dem des Sachfeldes zusammen (vgl. Fries 2016, S. 774).

## Bibliographie

**Bojanowski, Piotr, Edouard Grave, Armand Joulin und Tomas Mikolov.** 2016. „Enriching word vectors with subword information“. *arXiv preprint arXiv:1607.04606*.

**Ehrmantraut, Anton, Thora Hagen, Leonard Konle, und Fotis Jannidis.** 2021. „Type- and Token-Based Word Embeddings in the Digital Humanities“. *Proceedings of the Conference on Computational Humanities Research*, 16–38.

**Fries, Norbert.** 2016. „Wortfeld“. In Metzler Lexikon Sprache, hg. von Helmut Glück und Michael Rödel, Stuttgart: Metzler.

**Henrich, Verena und Erhard Hinrichs.** 2010. „GernE-diT - The GermaNet Editing Tool“. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC 2010)*, 2228–2235.

**Heuser, Ryan, und Long Le-Khac.** 2012. „A Quantitative Literary History of 2,958 Nineteenth-Century British Novels. The Semantic Cohort Method“. *Stanford Literary Lab Pamphlets*.

**Honnibal, Matthew, und Ines Montani.** 2017. „spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing“.

**Köper, Maximilian, und Sabine Schulte im Walde.** 2016. „Automatically Generated Norms of Abstractness, Arousal, Imageability and Valence for 350,000 German Lemmas“. In *Proceedings of the 10th Conference on Language Resources and Evaluation (LREC)*.

**Neuroth, Heike, Andrea Rapp, und Sibylle Söring** (Hgg). 2015. TextGrid. Von der Community - für die Community. Eine Virtuelle Forschungsumgebung für die Geisteswissenschaften. Glückstadt.

**Reuters, Hella** (Hg.). o.J. Projekt Gutenberg-DE, online unter: [www.projekt-gutenberg.org](http://www.projekt-gutenberg.org) [zuletzt aufgerufen am 14.12.2022].

**Underwood, Ted.** 2019. *Distant horizons. Digital Evidence and Literary Change*. Chicago: The University of Chicago Press.