

# Redewiedergabe in Hefromanen und Hochliteratur

**Brunner, Annelen**

brunner@ids-mannheim.de

Leibniz-Institut für Deutsche Sprache, Mannheim

**Jannidis, Fotis**

fotis.jannidis@uni-wuerzburg.de

Julius-Maximilians-Universität Würzburg, Deutschland

**Tu, Ngoc Duyen Tanja**

tu@ids-mannheim.de

Leibniz-Institut für Deutsche Sprache, Mannheim

**Weimer, Lukas**

lukas.weimer@uni-wuerzburg.de

Julius-Maximilians-Universität Würzburg, Deutschland

## Einführung

Die Art und Weise, wie die Rede und Gedanken einer Figur im Erzähltext eingebunden werden, ist einer der traditionellen Aspekte der Narratologie (vgl. z.B. Genette 2010; Martínez/Scheffel 2016). Die vorgestellte Studie untersucht die Anteile unterschiedlicher Redewiedergabeformen im Vergleich zwischen zwei Literaturtypen von gegensätzlichen Enden des Spektrums: Hochliteratur – definiert als Werke, die auf der Auswahlliste von Literaturpreisen standen – und Hefromanen, massenproduzierten Erzählwerken, die zumeist über den Zeitschriftenhandel vertrieben werden und früher abwertend als „Romane der Unterschicht“ (Nusser 1981) bezeichnet wurden. Unsere These ist, dass sich diese Literaturtypen hinsichtlich ihrer Erzählweise unterscheiden, und sich dies in den verwendeten Wiedergabeformen niederschlägt. Der Fokus der Untersuchung liegt auf der Dichotomie zwischen direkter und nicht-direkter Wiedergabe, die schon in der klassischen Rhetorik aufgemacht wurde (vgl. McHale 2014).

Die Studie geht von manuell annotierten Daten aus und evaluiert daran die Validität automatischer Annotationswerkzeuge, die im Anschluss eingesetzt werden, um die Menge des betrachteten Materials beträchtlich zu erweitern.

Zur Kontrastierung von Hefromanen und Hochliteratur mit quantitativen Methoden liegen bereits Studien vor, welche sich mit Fragen der sprachlichen und thematischen Komplexität beschäftigen (Jannidis/Konle/

Leinen 2019a/2019b). Das verwendete Annotationssystem sowie die Erkener wurden im Rahmen des Redewiedergabe-Projekts entwickelt (Brunner et al. 2019a/2019b).

## Voruntersuchung und Evaluation der automatischen Methoden

Für die Voruntersuchung wurden aus 22 Hochliteratur-Texten und 22 Hefromanen zufällige Textausschnitte von ca. 1000 Tokens gezogen. Da Hefromane typischerweise in Reihen mit unterschiedlichem Fokus erscheinen, betrachten wir auch das Verhalten dieser unterschiedlichen Hefroman-Genres. Die Hefroman-Ausschnitte wurden darum je zur Hälfte aus den Genres Liebesroman und Horrorroman gewählt. Die Texte wurden von zwei Erstannotatoren unabhängig voneinander bearbeitet. Eine dritte Person erstellte dann auf dieser Grundlage eine Konsensannotation, indem sie die beiden Annotationen verglich, Unstimmigkeiten bereinigte und wenn nötig offensichtliche Fehler korrigierte.

Das Annotationssystem (Brunner et al. 2019a) erfasst sowohl die Wiedergabe von Rede als auch von Gedanken und Geschriebenem. Es umfasst vier Haupttypen von Wiedergabe:

- direkt: *Er dachte*: „Ich habe Hunger.“
- frei-indirekt (‘erlebte Rede’): *Er war ratlos*. Wo sollte er jetzt etwas zu essen finden?
- indirekt: *Er sagte*, dass er Hunger habe.
- erzählt: Er sprach über das Mittagessen.

Frei-indirekte Wiedergabe ist im Folgenden ausgeschlossen, da sich die automatische Erkennung für diese Form als noch zu unzuverlässig erwiesen hat. Die Formen indirekte und erzählte Wiedergabe sind zu ‘nicht-direkt’ zusammengefasst. Diese Form umfasst damit sowohl die klassische indirekte Wiedergabe mit Rahmenformel und abhängiger Proposition als auch strukturell abweichende und häufig stärker zusammenfassende. Sie steht im Gegensatz zur direkten Wiedergabe insofern, als die Rede, Gedanken oder schriftliche Äußerung einer Figur in den Erzählertext integriert anstatt in einem Zitat klar davon abgesetzt wird.

Die automatischen Erkener beruhen auf DeepLearning.<sup>1</sup> Als Trainingsmaterial wurde hauptsächlich das Redewiedergabe-Korpus (Brunner et al. 2019b; verfügbar unter [github.com/redewiedergabe/corpus](https://github.com/redewiedergabe/corpus)) verwendet, welches historische Textdaten (19. bis frühes 20. Jahrhundert) und sowohl fiktionales als auch nicht-fiktionales Material umfasst. Da die in dieser Studie verwendeten Texte deutlich moderner sind (ca. 1950 bis Gegenwart), war es umso wichtiger, die Übertragbarkeit des Modells zu evaluieren. Es lagen speziell trainierte Erkener für jede der Formen direkte, indirekte und erzählte Wiedergabe vor, die unabhängig voneinander auf die Testdaten angewendet wurden. Überlagerungen von

Wiedergabetypen werden somit erkannt. Als ‚nicht-direkt‘ zählen Tokens, die entweder als Teil von indirekter oder von erzählter Wiedergabe erkannt wurden.

Um eine bessere Einschätzung der Erkennungswerte zu geben, ein paar Worte zur Vorkommenshäufigkeit der Redewiedergabetypen: Im den konsensannotierten Testdaten liegt der durchschnittliche Anteil von direkter Wiedergabe knapp unter 30% der Tokens (mit starken Schwankungen), von nicht-direkter bei ca. 15%.

Tabelle 1 zeigt die Übereinstimmungswerte zwischen den Erstannotatoren, um einen Eindruck zu vermitteln, wie verlässlich eine von Menschen durchgeführte Annotation wäre.

	F1	Precision	Recall	Fleiss' Kappa	prozentuale Anteile	
					durchschnittlicher absoluter Fehler	Standardabweichung des Fehlers
direkt						
Heftrömene	0,99	0,98	0,99	0,98	0,65	1,07
Hochliteratur	0,97	0,98	0,96	0,96	1,29	2,7
Alle Samples	0,98	0,98	0,98	0,97	0,97	2,0
nicht-direkt						
Heftrömene	0,77	0,78	0,77	0,74	1,96	1,46
Hochliteratur	0,78	0,78	0,78	0,74	2,60	2,64
Alle Samples	0,78	0,78	0,78	0,74	2,28	2,10

Tabelle 1: Übereinstimmung zwischen den Erstannotatoren. F1, Precision, Recall jeweils für die Kategorie direkt bzw. nicht-direkt, gerechnet auf Tokenbasis; prozentuale Anteile ebenfalls auf Tokenbasis.

Tabelle 2 zeigt nun für die Formen direkt und nicht-direkt die Übereinstimmungsquoten der automatischen Methoden im Vergleich zur Konsensannotation. Wenn man als Baseline einen Erkennen annimmt, der jedes Token mit 50% Wahrscheinlichkeit als Teil von Wiedergabe klassifiziert, käme man für die Testdaten (alle Samples) für direkt auf einen F1-Score von 0,36 (Precision: 0,28; Recall: 0,50), für nicht-direkt auf einen F1-Score von 0,23 (Precision: 0,17; Recall: 0,50), wobei die Einzelscores für Heftrömene vs. Hochliteratur bei direkt gleich wären, bei nicht-direkt etwas besser für Hochliteratur (F1: 0,25).

	F1	Precision	Recall	Fleiss' Kappa	prozentuale Anteile	
					durchschnittlicher absoluter Fehler	Standardabweichung des Fehlers
direkt						
Heftrömene	0,92	0,91	0,94	0,96	4,4	7,78
Hochliteratur	0,74	0,67	0,84	0,63	11,28	10,6
Alle Samples	0,83	0,78	0,89	0,76	7,84	9,91
nicht-direkt						
Heftrömene	0,75	0,76	0,73	0,75	1,69	1,59
Hochliteratur	0,66	0,73	0,61	0,60	4,23	3,26
Alle Samples	0,70	0,74	0,67	0,65	2,96	2,86

Tabelle 2: Auswertung der automatischen Methoden gegen die Konsens-Annotation.

Bei direkter Wiedergabe sind die Erkennungsraten der automatischen Methoden vor allem bei den Heftrömene gut, es gibt jedoch Schwankungen zwischen den Textausschnitten. Probleme treten insbesondere bei Ich-Perspektive in Kombination mit unmarkierter Wiedergabe auf, was in Hochliteratur häufiger vorkommt. Dennoch sind die mit dem maschinellen Erkennen erzielten Ergebnisse – gerade für solche Fälle – deutlich stabiler als eine Identifikation von direkter Wiedergabe anhand von Anführungszeichen gewesen wäre. Insgesamt neigt der Erkennen dazu, den Anteil von direkter Wiedergabe eher zu über- als zu unterschätzen. Der durchschnittliche absolute Fehler bei der Abschätzung der Anteile liegt im Schnitt bei ca. 10%.

Für die nicht-direkte Wiedergabe ist zu betonen, dass die Übereinstimmungsquote auch zwischen Menschen deutlich schlechter ist (vgl. Tabelle 1). Ursache ist, dass durch die stärkere Integration in den Erzähltext sowohl die genaue Abgrenzung als auch die Entscheidung, was als Wiedergabe zu werten ist, schwieriger sind. Die automatischen Methoden erreichen bei den Heftrömene fast gleiche Verlässlichkeit, während die Hochliteratur-Abschnitte sich wiederum als etwas schwieriger erweisen. Da die Anteile von nicht-direkt geringer sind und weniger Schwankungen unterliegen als die Anteile von direkt, ist auch der durchschnittliche absolute Fehler deutlich geringer (ca. 3%), wobei der Anteil von nicht-direkt eher unterschätzt wird.

Da für die Erzählweise eines Textes auch das Zusammenspiel der beiden Wiedergabetypen von Interesse ist, visualisieren wir die Textausschnitte der unterschiedlichen Untersuchungsgruppen in einem Scatterplot (Abb. 1).

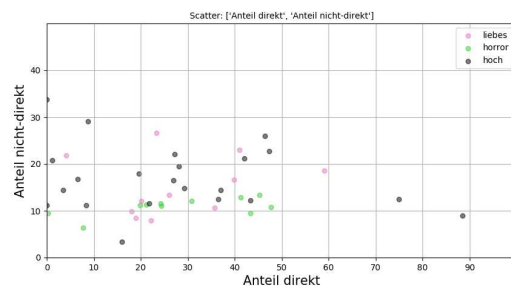


Abbildung 1: Scatterplot der 1000-Token-Samples auf Basis der manuellen Konsens-Annotation

In dieser Darstellung auf Basis der manuellen Konsens-Annotation lässt sich ein Trend der Horrorroman-Textausschnitte erkennen, mit niedrigen Werten sowohl in direkt als auch in nicht-direkt zusammenzuklustern, während Hochliteratur und Liebesroman stark gestreut

erscheinen. Mit einem Permutationstest ( $p=0,01$ ) (Koplenig 2019) lassen sich im Vergleich Heftroman vs. Hochliteratur allerdings auf keiner der beiden Dimensionen signifikante Unterschiede nachweisen. Bei dem Vergleich mit Genres sind lediglich die Abweichungen zwischen Hochliteratur und Horrorroman im Anteil nicht-direkter Wiedergabe signifikant. Legt man die automatisch annotierten Daten zugrunde, verschwindet auch diese Signifikanz.

## Erweiterung der Studie auf Volltexte

Im nächsten Schritt erweitern wir unser Untersuchungsmaterial stark. Das Korpus wurde aus Volltexten zusammengestellt, wobei diesmal die Unterschiede zwischen Hochliteratur und den einzelnen Genres in den Fokus gerückt wurden: 50 Hochliteratur-Texte wurden mit jeweils 50 Heftromanen aus vier unterschiedlichen Reihen kontrastiert, die unterschiedliche Genres repräsentieren (vgl. Tab. 3).

Gruppe	Anzahl Texte	Anzahl Autoren	Reihe	Tokens pro Text
Hochliteratur	50	50	-	35.310 - 332.568
Liebes	50	50	Julia Extra	19.608 - 42.908
Science-Fiction	50	11 (2-5 Texte pro Autor)	Perry Rhodan	28.365 - 34.904
Horror	50	4 (2-35 Texte pro Autor)	John Sinclair	27.920 - 34.838
Krimi	50	unbekannt; mind. 2	Jerry Cotton	29.212 - 50.171

Tabelle 3: Korpuszusammensetzung

Ein Ziel war, eine größtmögliche Diversität von Autoren zu erreichen, um zu verhindern, dass das Autorensignal die Gruppenzugehörigkeiten überlagert, die uns eigentlich interessieren. Problematisch war dies bei Horrorromanen, wo ein Autor die Reihe extrem dominiert und Krimis, bei denen so gut wie keine Autoreninformationen verfügbar waren. Es ist allerdings bekannt, dass die Reihe „Jerry Cotton“ von über 100 unterschiedlichen Autoren verfasst wurde (vgl. Karr 2019).<sup>2</sup> Da Heftromane üblicherweise unter Pseudonym veröffentlicht werden, ist die Autorenzuschreibung hier insgesamt mit Unsicherheit behaftet (vgl. Hügel 2001).

Die Texte wurden mit den automatischen Erkennern komplett annotiert. Da die Variation der Textlängen insbesondere in der Gruppe Hochliteratur stark ist, wurden die Texte in 1000-Token-Abschnitte zerlegt, für diese die Anteile von direkter und nicht-direkter Wiedergabe berechnet und die Ergebnisse anschließend für jeden Text gemittelt (analog zur standardisierten Type-Token-Ratio). Anders als bei den Testdaten zeigen sich bei der Auswertung nun klare Unterschiede in beiden Dimensionen: Der Anteil direkter Wiedergabe ist bei Hochliteratur geringer, während der Anteil nicht-direkter

Wiedergabe höher ist. Die Signifikanz dieser Unterschiede, wie auch viele Unterschiede zwischen den Genres, lassen sich mit dem Permutationstest mit  $p=0,01$  bestätigen (vgl. Abbildung 2 und 3).

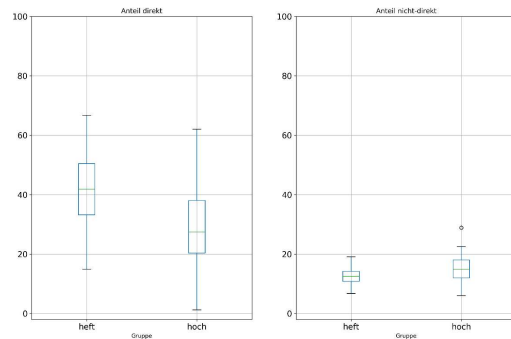


Abbildung 2: Boxplots Heftromane vs. Hochliteratur. Unterschiede sind signifikant mit  $p=0,01$ .

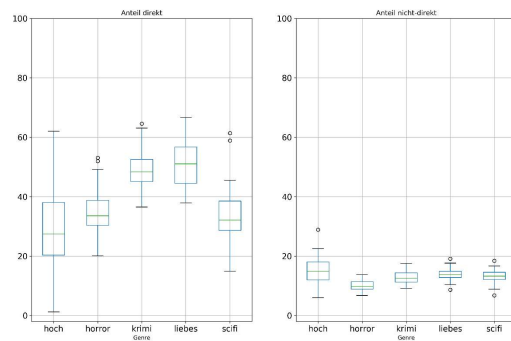


Abbildung 3: Boxplots Hochliteratur vs. Heftromangen-Genres: Signifikant mit  $p=0,01$  sind die Unterschiede hoch/krimi, hoch/horror, horror/krimi, horror/liebes (in beiden Dimensionen); hoch/liebes, scifi/liebes, scifi/krimi (nur in Anteil direkt); horror/scifi (nur in Anteil nicht-direkt).

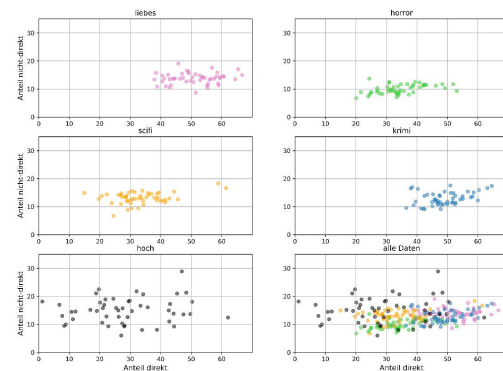


Abbildung 4: Scatterplots für die Volltexte, automatische Annotation

Bei der Betrachtung beider Dimensionen in Relation (Abb. 4) fällt sofort auf, dass die Hochliteratur-Texte eine deutliche Streuung aufweisen, während nicht nur die einzelnen Genres, sondern auch die Heftromane als Gruppe zusammenclustern. Angesichts der Tatsache, dass die Heftroman-Genres bewusst reglementierte Reihen sind, während die Hochliteratur-Gruppe nur dadurch definiert ist, dass die enthaltenen Werke als literarisch hochwertig eingeschätzt wurden, ist dieser Befund nicht erstaunlich. Es ist jedoch durchaus bemerkenswert, dass sich der Unterschied zwischen konventionalisiertem und individualistischem Erzählen auf der Dimension der Redewiedergabetypen so deutlich quantitativ nachweisen lässt.

Die Hochliteratur-Texte sind zudem die einzige Gruppe, in der ein ‚Übergewicht‘ an nicht-direkter im Gegensatz zu direkter Wiedergabe auftritt. Die Autoren sind also in der Art und Weise, wie sie Figurenstimmen in den Text einbinden, sowohl individualistischer als auch eher bereit, nicht das direkte Zitat zu wählen.

Innerhalb der Gruppe der Heftromane man kann für die Genres Liebesroman, Horrormoman und Krimi einen nahezu linearen Anstieg der beiden Wiedergabeformen in Relation zueinander beobachten, wobei der Anteil direkter Wiedergabe stets höher ist. Es differenziert sich recht klar das Horror-Genre mit einem insgesamt geringeren Wiedergabeanteil, während die ‚kommunikativeren‘ Genres Liebesroman und Krimi sich stark überlagern. Für diese beiden Genres lassen sich auch keine signifikanten Unterschiede nachweisen. Science-Fiction nimmt eine Zwischenstellung ein: Die Texte sind diverser und streuen ähnlich wie Hochliteratur, wenn auch nicht so extrem. Es ist das einzige Heftroman-Genre, für das sich auf keiner der beiden Dimensionen signifikante Unterschiede zu Hochliteratur nachweisen lassen. Dies passt zu Beobachtungen von Jannidis/Konle/Leinen (2019a), dass Science-Fiction unter den Heftroman-Genres eine Sonderstellung einnimmt und auch bei unterschiedlichen Komplexitätsmaßen wie standardisierter Type-Token-Ratio und Wortlänge höher abschneidet als die anderen Genres.

Warum zeigen sich diese interessanten Muster erst in den Volltextdaten und nicht in der Voruntersuchung? Die Erklärung ist, dass die Schwankungen in den Anteilen von Wiedergabe innerhalb eines Erzähltextes so stark sind, dass sie die beobachteten Trends überlagern. Abb. 5 zeigt einen Datenpunkt für jeden der 1000 Token-Abschnitte aus dem Untersuchungskorpus. Zwar werden in der Gesamtheit dieser Datenpunkte die gleichen Trends sichtbar wie in Abb. 4, doch wenn man – wie bei der Testauswertung – nur wenige zufällig gezogene 1000-Token-Abschnitte betrachtet, ist es unwahrscheinlich, dass sie erkennbar wären. Die Ausweitung auf mehr Material, die durch die Anwendung automatischer Methoden möglich wurde, führt

hier also zu einem Erkenntnisgewinn, der sonst nur mit extremem Annotationsaufwand möglich gewesen wäre.

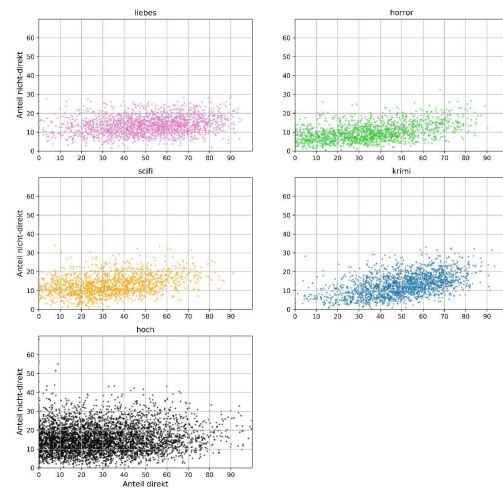


Abbildung 5: Scatterplots für die Volltexte (zerlegt in 1000-Token-Abschnitte), automatische Annotation.

Da die schlechteren Erkennungsraten des Direkte-Wiedergabe-Erkenner für Texte in der Ich-Perspektive bekannt sind, wurde für einen großen Teil der Texte die Erzählperspektive ermittelt. Die Durchmischung ist sowohl bei den Hochliteraturtexten als auch bei den Heftromanen gegeben und Texte beider Perspektiven platzieren sich an unterschiedlichen Stellen. Einzig der Bereich mit sehr niedrigem Anteil von direkter Wiedergabe (<17%) ist ausschließlich durch Texte in Er-Perspektive besetzt. Der Einfluss der Erzählperspektive ist ein Faktor, der in weiteren Untersuchungen genauer betrachtet werden sollte.

## Ausblick

Mit den vorhandenen Werkzeugen ist es denkbar, die Studien auf noch mehr Textmaterial auszuweiten und dabei auch weitere Genres von Heftromanen zu untersuchen. Zudem lässt sich die Methodik leicht auf Fragestellungen zu den Anteilen von Redewiedergabe in anderen Textgruppen übertragen, z.B. zwischen fiktionalem und nicht-fiktionalem Material oder im diachronen Vergleich. Wir arbeiten zudem im Redewiedergabe-Projekt weiter daran, unsere automatischen Erkenner zu verbessern, insbesondere auch den für freie-indirekte Wiedergabe (zum aktuellen Stand vgl. Brunner et al. 2019c). Die im Redewiedergabe-Projekt entwickelten Erkenner werden nach Abschluss des Projekts im Frühjahr 2020 der Forschungsgemeinschaft zur Verfügung gestellt werden, ebenso wie große Teile des verwendeten manuell annotierten Trainingsmaterials.

## Fußnoten

1. Wir verwenden eine BiLSTM-CRF-Architektur mit contextual string embeddings in Kombination mit FastText word embeddings (adaptiert von Akbik/Blythe/Vollgraf 2018). Eine detaillierte Beschreibung übersteigt den Rahmen dieses Beitrags, aber sie entspricht im Wesentlichen der Architektur in Brunner et al. 2019c.
2. Von den 50 Jerry-Cotton-Krimis konnten nur für 4 Autoreninformationen gefunden werden, von denen eine zudem unsicher ist. Gesichert kann man damit sagen, dass mindestens 2 unterschiedliche Autoren in der Gruppe Krimi vorliegen, vermutlich sind es jedoch deutlich mehr.

## Bibliographie

**Akbik, Alan / Blythe, Duncan / Vollgraf, Roland** (2018): "Contextual String Embeddings for Sequence Labeling", in: *Proceedings of the 27th International Conference on Computational Linguistics* 1638-1649.

**Brunner, Annelen / Weimer, Lukas / Engelberg, Stefan / Jannidis, Fotis / Tu, Ngoc Duyen Tanja** (2019a): *Annotationsrichtlinien des Projekts „Redewiedergabe. Eine literatur- und sprachwissenschaftliche Korpusanalyse“*. Zenodo. <http://doi.org/10.5281/zenodo.2634994> [letzter Zugriff 17. Dezember 2019].

**Brunner, Annelen / Weimer, Lukas / Tu, Ngoc Duyen Tanja / Engelberg, Stefan / Jannidis, Fotis** (2019b): "Das Redewiedergabe-Korpus. Eine neue Ressource", in: *Digital Humanities im deutschsprachigen Raum – Konferenzabstracts* 103-106.

**Brunner, Annelen / Tu, Ngoc Duyen Tanja / Weimer, Lukas / Jannidis, Fotis** (2019c): "Deep Learning for Free Indirect Representation", in: *Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019)* 241-245.

**Genette, Gérard** (2010): *Die Erzählung*. 3., durchges. und korrigierte Aufl. Paderborn: Fink.

**Hügel, Hans-Otto** (2001): "Kommunikative und ästhetische Funktion des Romanhefts", in: Leonhardt, Joachim-Felix / Ludwig, Hans-Werner / Schwarze, Dietrich / Straßner, Erich (eds.): *Medienwissenschaft. Ein Handbuch zur Entwicklung der Medien und Kommunikationsformen*. Berlin / New York: de Gruyter 1621-1631.

**Jannidis, Fotis / Konle, Leonard / Leinen, Peter** (2019a): "Makroanalytische Untersuchung von Hefromanen", in: *Digital Humanities im deutschsprachigen Raum – Konferenzabstracts* 167-173.

**Jannidis, Fotis / Konle, Leonard / Leinen, Peter** (2019b): "Thematic complexity", in: *Digital Humanities Conference* <https://dev.clariah.nl/files/dh2019/boa/0504.html> [letzter Zugriff 17. Dezember 2019].

**Karr, H.P.** (2019): "Cotton, Jerry", in: Ders. (ed.): *Lexikon der deutschen Krimi-Autoren – Internet Edition*

<http://www.krimilexikon.de/cotton.htm> [letzter Zugriff 17. Dezember 2019].

**Koplenig, A.** (2019): "A non-parametric significance test to compare corpora", in: *PLoS ONE* 14(9): e0222703 <https://doi.org/10.1371/journal.pone.0222703> [letzter Zugriff 17. Dezember 2019].

**Martínez, Matías / Scheffel, Michael** (2016): *Einführung in die Erzähltheorie*. 10. Auflage. München: C.H. Beck.

**McHale, Brian** (2014): "Speech Representation", in: Hühn, Peter / Pier, John / Schmid, Wolf / Schönert, Jörg (eds.): *The living handbook of narratology*. Hamburg: Hamburg University Press 434-446 <http://www.lhn.uni-hamburg.de/article/speech-representation> [letzter Zugriff 17. Dezember 2019].

**Nusser, Peter** (1981): *Romane für die Unterschicht. Groschenhefte und ihre Leser*. 5. Auflage. Stuttgart: Metzler.