

# Qualitative Genre-Profile und distinktive Wörter: Eine Studie zu Keynes in Subgenres des französischen Romans

**Röttgermann, Julia**

roettger@uni-trier.de  
Universität Trier, Deutschland  
ORCID: 0000-0002-1918-8117

**Du, Keli**

duk@uni-trier.de  
Universität Trier, Deutschland  
ORCID: 0000-0001-7800-0682

**Schöch, Christof**

schoech@uni-trier.de  
Universität Trier, Deutschland  
ORCID: 0000-0002-4557-2753

## Einleitung

„La différence n'est pensée que dans le jeu comparé de deux similitudes“

(Deleuze 2015, 238)

In *Differenz und Wiederholung* argumentiert Gilles Deleuze, dass Differenz nur in einem Spiel aus Ähnlichkeiten wahrgenommen werden kann, da Unterschiede erst durch den Vergleich mit Ähnlichem sichtbar werden. Mithilfe statistischer Keynes-Maße können distinktive Wörter in einer kontrastiven Betrachtung zwischen Textgruppen extrahiert werden. Die vorliegende Studie wurde im Rahmen des Projekts *Beyond Words*<sup>1</sup> durchgeführt, das statistische Maße der Distinktivität<sup>2</sup> auf Grundlage von Worthäufigkeit, Wortverteilung und Dispersion in den Blick nimmt und analysiert.<sup>3</sup>

In diesem Beitrag konzentrieren wir uns auf die Analyse der Untergattungen französischer Romane und versuchen, die Lücke zwischen qualitativen und quantitativen Aspekten in der Computational Literary Studies (CLS) zu schließen, indem wir ein Mapping zwischen qualitativen, expertenbasierten „Subgenre-Profilen“ und distinktiven Wörtern erstellen, die mit verschiedenen Distinktivitätsmaßen extrahiert wurden.

## Distinktivitätsmaße

In Computerlinguistik und CLS existiert eine zunehmend unübersichtliche Vielzahl an statistischen Maßen, um große Textmengen hinsichtlich distinktiver Wörter zu untersuchen. Dies ist teilweise begründet in fachlichen oder nationalen Traditionen, liegt aber auch teilweise an der Implementierung in Tools.<sup>4</sup> Wir unterteilen die Distinktivitätsmaße in drei Kategorien: häufigkeitsbasierte Maße, verteilungsbasierte Maße und dispersionsbasierte Maße. Für diese Studie haben wir aus jeder dieser drei Kategorien ein statistisches Maß ausgewählt, das sich in einer früheren Studie (Du, Dudar, und Schöch 2022) bewährt hat: logarithmisches Zeta (Burrows 2007; Schöch 2018), Welch's t-Test (Welch 1947) und Log-Likelihood-Ratio-Test (Dunning 1993).

**Der Log-Likelihood-Ratio-Test (LLR)** ist ein gut etabliertes, auf Häufigkeit basierendes Maß in der Computerlinguistik, hat jedoch den Nachteil, dass er auf der Gesamtwortzahl im Korpus basiert und häufige, aber ungleichmäßig verteilte Wörter bevorzugen kann.

**Welch's t-Test:** Ein verteilungsbasiertes Maß, das im Gegensatz zum klassischen t-Test keine Normalverteilung voraussetzt und daher für Sprachkorpora zuverlässiger ist.

**Zeta** ist ein primär dispersionsbasiertes Maß. Es wurde ursprünglich zur Autorschaftsattributions entwickelt und vergleicht die gleichmäßige Verteilung von Merkmalen in zwei Textsammlungen. Die logarithmische Zeta-Version wurde von Schöch et al. (2018) vorgeschlagen.

## Anwendungsgebiet: Der zeitgenössische französische Roman

Der französische populäre Roman hat eine lange Geschichte, die mindestens bis ins mittlere bis späte 19. Jahrhundert zurückreicht. Die Periode von 1860–1920 wird oft als „goldenes Zeitalter“ des populären französischen Romans angesehen, als Phänomene wie der „roman-feuilleton“ und der Fortsetzungsroman mit wiederkehrenden Protagonisten wie Rocombole und Rouletabille aufkamen. Frühere Studien zum Populärroman konzentrierten sich häufig auf das 19. und frühe 20. Jahrhundert (Angenot 1975; Olivier-Martin 2013).

Auch in der zweiten Hälfte des 20. Jahrhunderts existiert in Frankreich eine vielfältige Landschaft populärer Romane (Migozzi, 2005). Diese Romane werden von etablierten Verlagen (z.B. Harlequin, Fleuve noir, Éditions du Masque) in spezialisierten Sammlungen mit hohen Auflagen und einer klaren Zielgruppenorientierung herausgegeben. Einige Subgenres des Populärromans erhielten besondere Aufmerksamkeit, wie der „roman policier“ (Todorov 1971; Vannoncini 2002; Dubois 2005), der Science-Fiction-Roman (Slusser 1989; Thomas 1989; Millet und Labbé 2001; Baudou 2003; Mather und Rheault 2016) oder der Liebesro-

man / ‘roman sentimental’ (Helgorsky 1985; 1987; Constans 1999).



**Abb. 1** Vier Buchcover aus unserem Korpus zeitgenössischer französischer Romane in vier Subgenres: Kriminalroman, Science-Fiction-Roman, Sentimentalroman und 'littérature blanche' (Hochliteratur).

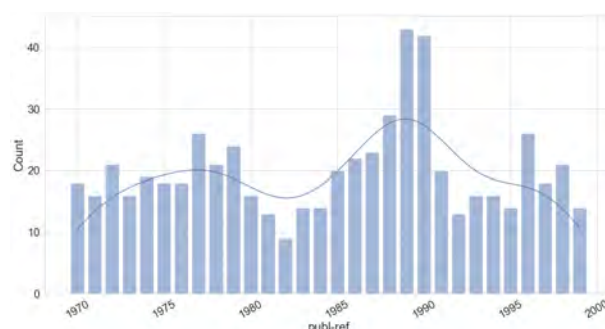
In jüngster Zeit wurde dem zeitgenössischen französischen populären Roman aus linguistischer Perspektive im Rahmen des Projekts ‚Phraséorom‘<sup>5</sup> und der Forschungsgruppe ‚Motifs‘ erneut Aufmerksamkeit gewidmet. Beide Gruppen entwickeln Techniken zur Extraktion lexico-grammatikalischer Muster aus Korpora, die nach Subgenres strukturiert sind (z.B. Gonon et al. 2017). Wir knüpfen an diese Forschung an und konzentrieren uns auf französische Populärromane und Hochliteratur aus der Zeit von 1970 bis 1999.

Serien von Kriminal- oder Science-Fiction-Romanen sind zudem Beispiele für serielles Erzählen, das auf eine lange Geschichte in der französischen Literatur zurückblicken kann.<sup>6</sup> Diese Subgenres enthalten wiederkehrende Formeln von Strukturelementen, typischen Protagonisten, Erzählmustern, wiederkehrenden Schauplätzen und charakteristischen sprachlichen Mustern. Aufgrund dieser Muster und Schemata sind sie besonders geeignet als Untersuchungsmaterial zur Extraktion distinktiver Wörter pro Gattung und pro Keyness-Maß. Die zu erwartenden Muster an Themen, Figuren, Schauplätzen, Tonalitäten oder sprachlichen Spezifika sind eingegrenzter als in einem breit angelegten Korpus französischer Literatur.

Um gemeinsame Merkmale von Kriminalromanen, Science-Fiction-Romanen, Sentimentalromanen und Hochliteratur zu identifizieren, haben wir zunächst eine Liste typischer Elemente wie Themen, Figuren, sprachliche Muster, Tonalität und Erzählform durch Sichtung relevanter Sekundärliteratur qualitativ zusammengestellt. Diese Profile dienten als Grundlage für die Kategorisierung distinktiver Wörter pro Subgenre, die von den Distinktivitätsmaßen identifiziert wurden.

## Daten

Es wurde ein Korpus französischer Romane der 1970er, 1980er und 1990er Jahre erstellt, welches vier Gattungen enthält: Science-Fiction, Sentimentalroman, Kriminalroman und Hochliteratur, die in Frankreich sogenannte ‚littérature blanche‘.<sup>7</sup>



**Abb. 2** Anzahl an Romanen pro Jahr im Korpus.

Das Korpus zeitgenössischer französischer Literatur enthält 33 Millionen Tokens und umfasst 600 Romane, gleichmäßig verteilt auf vier Subgenres (Sentimentalromane, Kriminalromane, Science-Fiction-Romane, Hochliteratur) und drei Zeitperioden (1970er, 1980er, 1990er Jahre). Somit besteht das Korpus aus 50 Romanen pro Jahrzehnt und Subgenre. Die Jahrzehnte sind gleichmäßig vertreten, allerdings ist ein Höhepunkt der Veröffentlichungen in den Jahren 1989–1990 zu verzeichnen (Abb. 2).

Da alle Romane urheberrechtlich geschützt sind, wurde das Korpus in Form von „abgeleiteten Textformaten“ (Schöch et al. 2020) veröffentlicht, einem Format, das für bestimmte computergestützte Analysen geeignet ist, aber für Menschen unlesbar bleibt. Diese Strategie soll Transparenz und Reproduzierbarkeit der Studie ermöglichen.

## Methoden

Mit dem Ziel, distinktive Wörter pro Subgenre mit drei verschiedenen Distinktivitätsmaßen zu extrahieren, vergleichen wir jedes Subgenre mit allen anderen, z. B. Science-Fiction vs. den Rest. Dazu werden 150 Romane eines Subgenres mit 450 Romanen der übrigen Subgenres verglichen.

Da Französisch eine flektierte Sprache ist, werden alle Texte mit SpaCy lemmatisiert (Montani et al. 2023). Anschließend werden sie mit dem Python-Paket pydistinto (Du et al. 2022) verarbeitet. Die drei angewendeten Distinktivitätsmaße sind Zeta, Welch und LLR. Für die Berechnung von Zeta werden die Romane in 5000-Wörter-Segmente unterteilt. Aus jedem Test erhalten wir drei Listen mit distinktiven Wörtern und wählen die Top 50 Wörter für den Abgleich mit den qualitativen Subgenre-Profilen aus.

Unsere Erwartung ist dabei, dass die extrahierten distinktiven Wortlisten pro Subgenre in mehr oder weniger ausgeprägtem Umfang Wörter enthalten, die sich auf die thematischen Konzepte, Sprachmuster, Figuren, Schauplätze, Tonalität, Erzählform oder Erzählstruktur aus den qualitativen Subgenre-Profilen beziehen lassen und so eine qualitative Evaluation der Maße erlauben.

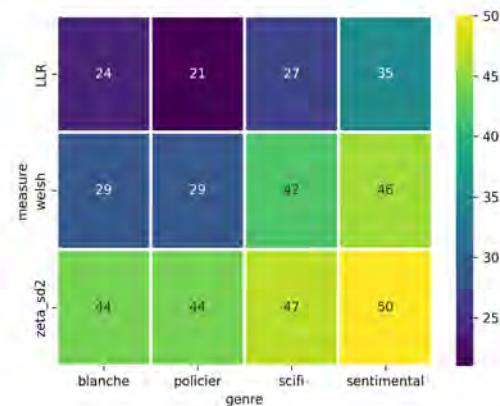
Genre-Profil	log. Zeta	LLR	Welch's t-test
Language Patterns: The linguistic landscape of French detective novels is characterized by sociolects, informal register and the simulation of direct speech, which bring authenticity to the narrative.	cop (flic), screw (foutre), guy (type), fella (gars), guy (mec), dumb (con), stuff (truc), shit (merde), eh (hein), buddy (copain), job (boulot), face (gueule), kid (môme)	say (dire), cop (flic), guy (mec)	cop (flic), say (dire), yes (oui), screw (foutre), guy (type), dumb (con), stuff (truc), shit (merde), ah (ah)
[Anteil kategorisierbarer Wörter]	13 / 50 words	3/50 words	9/50 words

**Tab. 1** Beispiel-Zeile: Matching von Genre-Profilen und extrahierten, distinktiven Wörtern. Die englische Übersetzung ist zur besseren Verständlichkeit hinzugefügt.<sup>8</sup>

Das Ziel ist es, den Anteil der interpretierbaren distinktiven Wörter pro Maß zu quantifizieren. Wörter, die nicht zu den Subgenre-Profilen passen, werden als "unerwartet" klassifiziert.<sup>9</sup>

Was verstehen wir im hier skizzierten Kontext unter Interpretierbarkeit? In dieser Studie gehen wir über die Klassifizierung von Keynes-Maßnahmen allein nach ihrer Leistung hinaus. Unter dem breiteren Paradigma von explainability (Erklärbarkeit) und interpretability (Interpretierbarkeit) von algorithmischen Methoden (u.a. Benois-Pineau et al. 2023), untersuchen wir die Ergebnislisten der Extraktion mit Keynes-Maßen. Insbesondere gleichen wir die Top 50 der distinktiven Wörter aus diesen Listen mit den von Fachwissenschaftler:innen erstellten Genreprofilen ab. Wir definieren dabei Interpretierbarkeit auf der Ebene der einzelnen Wörter: Ein distinktives Wort wird als interpretierbar definiert, wenn menschliche Annotator:innen es einem deskriptiven Genreprofil zuordnen kann. In einem ähnlichen Sinne haben Chang et al. (2009) einen Ansatz entwickelt, bei dem menschliche Annotator:innen die Aufgabe haben, aus einer Reihe von Topic Modeling Ergebnissen "the odd one out" (das nicht passende Wort innerhalb der Wortliste) zu identifizieren, als Operationalisierung der Interpretierbarkeit von Topic Modeling Ergebnissen. Bei der Beurteilung der Interpretierbarkeit lautet die übergreifende Frage außerdem: "Can we trust the model?" (Masís 2023, 6)? In unserem Fall lautet diese Frage demnach analog: Können wir dem Keynes-Maß (für diese Aufgabe im genannten Setting) vertrauen? Je höher der Anteil interpretierbarer Wörter ist, als desto zuverlässiger kann das Maß für ähnliche Aufgaben angesehen werden.

## Ergebnisse



**Abb. 3** Heatmap der zugeordneten, interpretierbaren Wörter pro Keynes-Maß und Subgenre.

**Logarithmisches Zeta** Bei der Analyse der Kriminalromane mit logarithmischem Zeta im Matching mit den qualitativen Profilen wurden distinktive Wörter wie *meurtre*, *crime*, *tuer*, *police*, *inspecteur*, *commissaire*, *victime*, *suspect* (‚Mord‘, ‚Verbrechen‘, ‚töten‘, ‚Polizei‘, ‚Inspektor‘, ‚Kommissar‘, ‚Opfer‘, ‚Verdächtiger‘), sowie distinktive Wörter, die auf Soziolekt und direkte Rede verweisen, festgestellt. Bei Science-Fiction-Romanen wurden typische thematische Konzepte wie *humain*, *cerveau*, *contrôle*, *puissance*, *détruire* (Mensch, Gehirn, Kontrolle, Macht, zerstören) identifiziert. Sentimentalromane zeigten Wörter wie *embrasser*, *ravir*, *bouleverser*, *émotion*, *sentiment* (küssen, entzücken, aufregen, Emotion, Gefühl). Bei Hochliteratur fanden sich Wörter wie *guerre*, *enfance*, *famille*, *France* (Krieg, Kindheit, Familie, Frankreich), die sich der Kategorie thematischer Statements zuordnen ließen. Insgesamt konnten 92,5% (185 aus 200) der Top-Keywords einem Aspekt des Subgenres zugeordnet werden (siehe Abb. 3).

**Welch's t-Test** identifizierte bei Kriminalromanen Wörter wie *enquête*, *téléphone*, *police* (Ermittlung, Telefon, Polizei). Weitere thematische Begriffe wie *meurtre* (Mord), *affaire* (Affäre), *coup* (Schuss), *crime* (Verbrechen), *assassiner* (ermorden), *tuer* (töten), *tirer* (schießen) sowie Begriffe des urbanen Raums wie *voiture* (Auto) und *bureau* (Büro) entsprechen den Erwartungen des Detektivgenres. Für Science-Fiction wurden Wörter wie *humain* (menschlich), *atteindre* (erreichen), *vaisseau* (Raumschiff), *autre* (andere), *contrôle* (Kontrolle), *nouveau* (neu), *cerveau* (Gehirn), *conscience* (Bewusstsein), *contact* (Kontakt), *pouvoir* (Macht), *puissance* (Kraft) extrahiert. Zudem zeigten sich zahlreiche Treffer im Bereich des technischen Vokabulars. Adjektive, die Größen beschreiben (z.B. *huge*, *gigantic*, *tiny*), sind laut Ted Underwood (2019) verlässliche Indikatoren für Science-Fiction. Entsprechende Wörter wie *centaine* (Hundert), *mètre* (Meter), *nombre* (Zahl) und *nombreux* (zahlreich) wurden im vorliegenden Korpus ebenfalls

beobachtet. Sentimentalromane enthalten distinktive Wörter wie *murmurer*, *lèvre*, *aimer*, *cœur*, *soupir* (dt. flüstern, Lippe, lieben, Herz, Seufzer). Die charakteristische Erzählform, für die anderen Subgenres eine schwer greifbare Kategorie, zeigte sich im sentimentalromanen besonders deutlich durch zahlreiche Kommunikationsverben und Anzeichen direkter Rede. Romane der Hochliteratur wiesen Wörter wie *vie*, *fin*, *enfance*, *misère* (Leben, Ende, Kindheit, Unglück) auf. Des weiteren wurde als distinktives Wort *écrire* (schreiben) für Hochliteratur identifiziert, ein Motiv, das ebenfalls von Forschenden des Phraséorom-Projekts für das Subgenre der *Littérature blanche* beschrieben ist (Gonon et al. 2018). Insgesamt konnten 73% (146 aus 200) der Top-Keywords einem Aspekt des Subgenres zugeordnet werden (siehe Abb. 3).

**Log-Likelihood-Ratio-Test** LLR zeigte bei Kriminalromanen distinktive Wörter zu typischen Figuren wie *inspecteur*, *commissaire*, *policier*, *shérif* (dt. Inspektor, Kommissar, Polizist, Sheriff) oder das Pronomen *je* (ich), welches auf eine gehäuft auftretende introspektive Erzählhaltung verweist. Bei Science-Fiction wurden nur wenige zu den Genreprofilen passende Wörter wie *vaisseau*, *humain*, *planète*, *espace* (dt. Raumschiff, Mensch, Planet, Weltraum) in den Kategorien der typischen Themen und Orte extrahiert. Wortlisten aus Kriminalromanen und Hochliteratur zeigten zudem größere Anteile an Wörtern der Kategorie 'unerwartet'. Wir können (Brezina 2018, 84–85) in dieser Hinsicht zustimmen, dass es zunehmend Belege dafür gibt, dass Log-Likelihood Ratio im Korpusvergleich viel zu viele „false hits“ erzeugt.<sup>10</sup> Insgesamt konnten nur 53,5% (107 aus 200) der Top-Keywords einem Aspekt des Subgenres zugeordnet werden (siehe Abb. 3).

## Fazit

Logarithmisches Zeta erwies sich beim Matching-Prozess der Wortlisten mit den Genreprofilen als am Besten geeignet zur Extraktion distinktiver Wörter, gefolgt von Welch's t-Test. Der Log-Likelihood-Ratio-Test zeigte im untersuchten Setting die schwächste Leistung, was überrascht, da das Maß in zahlreichen Korpusanalysetools wie Antconc oder Wordsmith Tools implementiert ist. Dieses Ergebnis wirft Fragen nach den standardmäßig implementierten Keyness-Maßen in DH-Tools auf und deutet darauf hin, dass eine kritische Überprüfung in Betracht gezogen werden sollte, zumindest in Szenarien, die mit narrativer Prosa arbeiten.

## Daten und Code

- Korpus und Metadaten: <https://github.com/Zeta-and-Company/dtf600>, DOI: 10.5281/zenodo.10853581.
- Code und Forschungsdaten <https://github.com/Zeta-and-Company/expertise-statistics>, DOI: 10.5281/zenodo.10853663.

- Pydistinto: <https://github.com/Zeta-and-Company/py-distinto>, DOI: 10.5281/zenodo.6517683.

## Fußnoten

1. Projekt Beyond Words: <https://tcdh.uni-trier.de/en/projekt/beyond-words> (zugegriffen: 27. November 2024).
2. Keyness-Maß und Distinktivitätsmaß werden in der Untersuchung als synonyme Begriffe verwendet.
3. Die statistische Auswertung von Worthäufigkeiten kann in der Sprachwissenschaft auf eine lange Geschichte zurückblicken; ein frühes Beispiel einer solchen nicht-computationellen Studie ist Yules „De Imitatione Christi“ (1944). Ziel solcher Studien war häufig die Bestimmung der Autorschaft, heute auch unter dem Namen ‚Stilometrie‘ eine beliebte Methode der Digital Humanities (Burrows 2007; Craig und Kinney 2009; Kestemont et al. 2018; Savoy 2020; Rebora 2023, 203).
4. Ein Beispiel ist das Maß der ‘spécificité’, das vor allem in französischen Tools wie TXM (Heiden et al. 2016) oder IRaMuTeQ (Ratinaud 2008) implementiert ist.
5. Phraséorom : la phraséologie du roman contemporain français, anglais et allemand, ANR-15-FRAL-0009 <https://phraséorom.univ-grenoble-alpes.fr> (zugegriffen: 27. November 2024).
6. Beispiele für seriell Erzählen finden sich im französischen Roman des 19. Jahrhunderts, zum Beispiel im Rougon-Macquart-Zyklus von Émile Zola oder in der *Comédie humaine* von Balzac.
7. Der Name ‚littérature blanche‘ spielt auf das häufig weiße Buchcover der *Edition de Gallimard* an. Der Verlag hat auch eine langjährige, prestigeträchtige Reihe ‘Collection Blanche’.
8. Die vollständige Tabelle ist hier einsehbar: GitHub Repository <https://github.com/Zeta-and-Company/expertise-statistics>, DOI: 10.5281/zenodo.10853663 (zugegriffen: 27. November 2024).
9. Der Abgleich zwischen Wortlisten und Subgenre-Profilen wurde unabhängig von drei Annotator:innen durchgeführt. Das mittlere Inter-Annotator-Agreement (Krippendorff's alpha) betrug  $\kappa=0,083$  für die Kategorisierung und  $\kappa=0,139$  für die Interpretierbarkeit der Wörter.
10. Siehe auch (Brezina / Meyerhoff 2014) oder (Tranchese 2023, 99).

## Bibliographie

- Angenot, Marc.** 1975. *Le roman populaire: recherches en paralittérature*. Montréal: Presses de l'université du Québec.
- Baudou, Jacques.** 2003. *La Science-fiction*. 1e édition. Que sais-je? Paris: Presses Universitaires de France - PUF.
- Benois-Pineau, Jenny, Romain Bourqui, Dragutin Petkovic und Georges Quenot.** 2023. *Explainable Deep Learning AI: Methods and Challenges*. Elsevier.

- Brezina, Vaclav.** 2018. *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press.
- Burrows, John.** 2007. „All the Way Through: Testing for Authorship in Different Frequency Strata“. *Literary and Linguistic Computing* 22 (1): 27–47. <https://doi.org/10.1093/lc/fqi067> (zugegriffen: 27. November 2024).
- Chang, Jonathan, Sean Gerrish, Chong Wang, Jordan Boyd-Graber und David M. Blei.** 2009. „Reading Tea Leaves: How Humans Interpret Topic Models“. In *Advances in Neural Information Processing Systems* 22. <http://www.umiacs.umd.edu/~jbg/docs/nips2009-rtlpres.pdf> (zugegriffen: 27. November 2024).
- Constans, Ellen.** 1999. *Parlez-moi d'amour: le roman sentimental: des romans grecs aux collections de l'an 2000*. Presses Univ. Limoges.
- Craig, Hugh und Arthur F. Kinney,** Hrsg. 2009. *Shakespeare, Computers, and the Mystery of Authorship*. 1. Aufl. Cambridge University Press.
- Deleuze, Gilles.** 2015. *Différence et répétition*. Paris: Humensis.
- Du, Keli, Julia Dudar, Cora Rok und Christof Schöch.** 2022. „Kontrastive Textanalyse mit pydistinto - Ein Python-Paket zur Nutzung unterschiedlicher Distinktivitätsmaße“. <https://doi.org/10.5281/zenodo.6327967> (zugegriffen: 27. November 2024).
- Du, Keli, Julia Dudar und Christof Schöch.** 2022. „Evaluation of Measures of Distinctiveness: Classification of Literary Texts on the Basis of Distinctive Words“. *Journal of Computational Literary Studies* 1 (1). <https://doi.org/10.48694/JCLS.102> (zugegriffen: 27. November 2024).
- Dubois, Jacques.** 2005. *Le roman policier ou la modernité*. Armand Colin.
- Dunning, Ted.** 1993. „Accurate Methods for the Statistics of Surprise and Coincidence“. *Computational Linguistics* 19 (1): 14. <http://aclweb.org/anthology/J93-1003> (zugegriffen: 27. November 2024).
- Gonon, Laetitia, Vannina Goossens, Olivier Kraif, Iva Novakova und Julie Sorba.** 2018. „Motifs textuels spécifiques au genre policier et à la littérature blanche“, hg. von F. Neveu, B. Harmegnies, L. Hriba, und S. Prévost. *SHS Web of Conferences* 46:06007. <https://doi.org/10.1051/shsconf/20184606007> (zugegriffen: 27. November 2024).
- Gonon, Laetitia, Vannina Goossens, Iva Novakova und Iva Novakova.** 2017. „Les phraséologismes spécifiques à deux sous-genres de la paralittérature : le roman sentimental et le roman policier“. In *Colloque Phraséologie française*. Arras: Université d'Artois.
- Helgorsky, Françoise.** 1985. „Harlequin ou la quête du grand amour“. *Communication et langages* 63 (1): 83–98. <https://doi.org/10.3406/colan.1985.1666> (zugegriffen: 27. November 2024).
- Helgorsky, Françoise.** 1987. „Harlequin: l'unité dans la diversité et vice-versa“. *Pratiques* 54 (1): 5–19. <https://doi.org/10.3406/prati.1987.1434> (zugegriffen: 27. November 2024).
- Kestemont, M., Michael Tschuggnall, E. Stamatatos, Walter Daelemans, Günther Specht, Benno Stein und Martin Potthast.** 2018. „Overview of the Author Identification Task at PAN-2018: Cross-domain Authorship Attribution and Style Change Detection“. In *Conference and Labs of the Evaluation Forum*.
- Mather, Philippe und Sylvain Rheault.** 2016. *Rediscovering French Science-Fiction in Literature, Film and Comics: From Cyrano to Barbarella*. Cambridge: Cambridge Scholars Publishing.
- Millet, Gilbert und Denis Labbé.** 2001. *La science-fiction*. Collection Sujets. Paris: Belin.
- Montani, Ines, Matthew Honnibal, Matthew Honnibal, Adriane Boyd, Sofie Van Landeghem und Henning Peters.** 2023. „explosion/spaCy: v3.7.2: Fixes for APIs and requirements“. Zenodo. <https://doi.org/10.5281/zenodo.10009823> (zugegriffen: 27. November 2024).
- Olivier-Martin, Yves.** 2013. *Histoire du roman populaire en France de 1840 à 1980*. Paris: Albin Michel.
- Ratinaud, Pierre.** 2008. „IRaMuTeQ“. Website. Iramuteq. 2008. <http://www.iramuteq.org> (zugegriffen: 27. November 2024).
- Rebora, Simone.** 2023. „GPT-3 vs. Delta. Applying Stylometry to Large Language Models“. In *La memoria digitale: forme del testo e organizzazione della conoscenza. Atti del XII Convegno Annuale AIUCD*, 292–97. <https://iris.univr.it/handle/11562/1115882> (zugegriffen: 27. November 2024).
- Savoy, Jacques.** 2020. „Elena Ferrante: A Case Study in Authorship Attribution“. In *Machine Learning Methods for Stylometry*, von Jacques Savoy, 191–210. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-53360-1\\_8](https://doi.org/10.1007/978-3-030-53360-1_8) (zugegriffen: 27. November 2024).
- Schöch, Christof.** 2018. „Zeta für die kontrastive Analyse literarischer Texte. Theorie, Implementierung, Fallstudie“. In *Quantitative Ansätze in den Literatur- und Geisteswissenschaften. Systematische und historische Perspektiven*, hg. von Toni Bernhart, Sandra Richter, Marcus Lepper, Marcus Willand, und Andrea Albrecht, 77–94. Berlin: De Gruyter. <https://www.degruyter.com/view/books/9783110523300/9783110523300-783110523300-004.xml> (zugegriffen: 27. November 2024).
- Schöch, Christof, Frédéric Döhl, Achim Rettinger, Evelyn Gius, Peer Trilcke, Peter Leinen, Fotis Jannidis, Maria Hinzmann und Jörg Röpke.** 2020. „Abgeleitete Textformate: Text und Data Mining mit urheberrechtlich geschützten Textbeständen“. *Zeitschrift für digitale Geisteswissenschaften (ZfdG)* 5. [http://dx.doi.org/10.17175/2020\\_006](http://dx.doi.org/10.17175/2020_006) (zugegriffen: 27. November 2024).
- Slusser, George.** 1989. „Science Fiction in France: An Introduction“. *Science Fiction Studies* 16 (3): 251–53. <https://www.jstor.org/stable/4239952> (zugegriffen: 27. November 2024).
- Thomas, Pascal J.** 1989. „The Current State of Science Fiction in France (L'état actuel de la science-fiction en France)“. *Science Fiction Studies* 16 (3): 298–306. <https://doi.org/10.1086/sfs.16.3.298>



[www.jstor.org/stable/4239955](http://www.jstor.org/stable/4239955) (zugegriffen: 27. November 2024).

**Todorov, Tzvetan.** 1971. „Typologie du roman policier“. In *Poétique de la prose*, 55–65. Paris: Seuil.

**Tranchese, Alessia.** 2023. *From Fritzl to #metoo: Twelve Years of Rape Coverage in the British Press*. Portsmouth: Springer Nature. <https://doi.org/10.1007/978-3-031-09353-1> (zugegriffen: 27. November 2024).

**Underwood, Ted.** 2019. *Distant horizons: digital evidence and literary change*. Chicago: The University of Chicago Press.

**Vanoncini, André.** 2002. *Le roman policier*. 3. éd., mise à jour. Que sais-je? 1623. Paris: Presses Univ. de France.

**Welch, Bernard Lewis.** 1947. „The Generalization of Student's Problem When Several Different Population Variances Are Involved“. *Biometrika* 34 (1–2): 28–35. <https://doi.org/10.1093/biomet/34.1-2.28>.

**Yule, C. Udny.** 1944. *The statistical study of literary vocabulary*. Cambridge: Cambridge University Press. <https://www.mpi.nl/publications/item2407784/statistical-study-literary-vocabulary> (zugegriffen: 27. November 2024).