

Multimodale Stilometrie: Herausforderungen und Potenzial kombinatorischer Bild- und Textanalysen am Beispiel Comics

Dunst, Alexander

alexander.dunst@gmail.com
Universität Paderborn, Deutschland

Hartel, Rita

rst@upb.de
Universität Paderborn, Deutschland

Einleitung

Stilometrische Untersuchungen blicken auf eine lange Tradition in der Literaturwissenschaft zurück (Holmes). Im Gegensatz dazu befinden sich quantitative Untersuchungen des Stils visueller Kunst und multimodaler Medien in einem frühen Experimentierstadium, das von explorativen Untersuchungen, Methodenentwicklung und -Adaption geprägt ist. Auch hier sind Fortschritte erkennbar, etwa in der digitalen Kunstgeschichte, der Filmwissenschaft und in der Comicforschung (Manovich, Douglas & Zepel; Qui, Taeb & Hughes; Baxter, Khitrova & Tsivian; Cutting et al.; Dunst & Hartel 2018a). Dabei konzentriert sich die stilistische Klassifikation bisher entweder auf visuelle oder sprachliche Kanäle. Beispielhaft zu nennen sind die Analyse der historischen Entwicklung von Filmfarben bei Barbara Flückiger oder Ben Schmidts thematische Untersuchungen populärer TV-Serien (Flückiger; Schmidt). Während dieser monomodale Fokus bei visueller Kunst der oftmaligen Dominanz der Bildebene geschuldet ist, so sind dafür bei multimodalen Medien wie Film, Fernsehen, Computerspielen oder Comics andere Gründe ausschlaggebend. Mit dem Topic Modeling oder der Textstilometrie stehen Methoden zur Verfügung, die auf sprachlichen Daten basieren und in der digitalen Literaturwissenschaft laufend weiterentwickelt werden. Visuelle Stilometrie, obwohl weit weniger ausgereift, kann auf die erwähnten Arbeiten aus der Kunstgeschichte und empirischen Filmwissenschaft zurückgreifen.

Auch technische Hürden tragen zu monomodalen Analysen bei: je nach Medium ist die digitale Erschließung und Analyse von Informationskanälen mit erheblichen Schwierigkeiten verbunden, etwa die Spracherkennung von Filmdialogen, die automatische Erkennung von handschriftlichen Texten in Comics

oder die computergestützte Verarbeitung großer Mengen an Bildmaterial. Die Kombination unterschiedlicher Informationskanäle in visuellen Medien führt jedoch unweigerlich zur Frage, inwieweit Stil durch die Analyse einzelner Modi erfasst werden kann. Auch thematische Untersuchungen setzen sich dem Vorwurf aus, komplexe Medien auf zu schmaler Datenbasis zu interpretieren, wenn Filmanalysen alleine auf Untertiteln oder Drehbüchern basieren. Im Gegenzug verbindet sich mit der Einbeziehung mehrerer Informationskanäle in stilometrische Untersuchungen die Hoffnung, multimodale Medien vollständiger beschreiben und einzelne Autoren, Genres oder Epochen genauer voneinander unterscheiden zu können. Im Folgenden werden erste Untersuchungen vorgestellt, die auf Basis eines Corpus an englischsprachigen Comicbüchern – so genannten „Graphic Novels“ – visuelle und Textstilometrie kombinieren (Dunst, Hartel & Laubrock).

Vorarbeiten und Herausforderungen

Wie bereits dokumentiert, haben wir auf Basis kunsthistorischer und filmwissenschaftlicher Vorarbeiten eine visuelle Stilometrie für Comicbücher entwickelt, die zwischen einzelnen Genres, Autoren und Publikationsformen unterscheiden und diese auf repräsentativer Datenbasis stilistisch beschreiben kann (Dunst & Hartel 2018b). Die relativ geringe Datenbasis, die mit der Analyse eines kulturellen Nischenproduktes einhergeht, bedeutete jedoch, dass nicht in allen Fällen signifikante Ergebnisse erzielt werden konnten. Insbesondere die Entwicklung einzelner Gattungen innerhalb eines Mediums lässt sich im historischen Verlauf nicht signifikant belegen. Auch nationale Traditionen konnten nicht immer stilistisch voneinander unterschieden werden, selbst wo sich diese für den Betrachter deutlich voneinander unterscheiden. Abbildung 1 zeigt, dass die von uns verwendeten visuellen Maße den japanischen Manga-Autor Osamu Tezuka stilistisch nicht von anglo-amerikanischen Werken abgrenzen konnten (Dunst & Hartel 2018a).

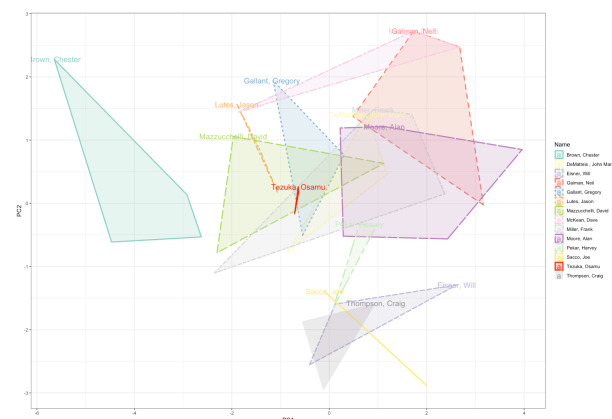


Abbildung 1. Visueller Stil bei Comics-Autoren

In beiden Fällen – also sowohl bei Autor- als auch bei Genreunterscheidungen – liegt es nahe, dass die kombinatorische Analyse von visueller und Textstilistik die Wahrscheinlichkeit erfolgreicher Unterscheidungen erhöhen würde. Allerdings stellen diese zusätzlichen Dimensionen eine multimodale Stilometrie vor methodische Herausforderungen. Wird eine große Zahl an Maßen für die Klassifikation herangezogen, so erschwert dies die qualitative Interpretation der Ergebnisse. Zwar lässt sich mit Hilfe einer Principal Component Analysis (PCA) darstellen, ob sich Genres oder Autoren signifikant unterscheiden. Je mehr Dimensionen in die PCA einfließen, desto schwerer fällt allerdings die Aussage, auf welchen Maßen diese Unterschiede fußen. Hinzu kommen wie erwähnt technische Hürden: abgesehen von den Angeboten bekannter Softwaregiganten führen automatische Spracherkennungssysteme noch zu relativ schlechten Resultaten. Ähnlich liegt der Fall bei Comics, deren Texte auf Handschriften basieren. Erst seit kurzem können diese Texte mit Hilfe automatischer Erkennungssysteme, die auf „Deep Learning“ basieren, zugänglich gemacht werden. Dennoch liegen Fehlerraten weit über jenen, die die Basis der meisten literaturwissenschaftlichen Analysen bilden. Der nächste Abschnitt stellt eine Methode vor, die dennoch die Verwendung einfacher Textmaße für die multimodale Stilometrie ermöglicht.

Datenbasis & Methode

Die Analysen basieren auf dem ersten repräsentativen Corpus englischsprachiger Comicbücher, von uns „Graphic Narrative Corpus“ (GNC) genannt (Dunst, Hartel & Laubrock). Wie in früheren Arbeiten beschrieben (Hartel & Dunst), nutzen wir die Bag Error Rate (BER) für eine Abschätzung, ob die Qualität der erkannten Texte ausreichend gut ist, um diese für die Textanalysen heranzuziehen. Hierzu haben wir als Gold Standard rund 10% der Seiten einer Graphic Novel manuell annotiert. Wir betrachten in unseren Analysen die Multi-Menge (oft als „Bag“ bezeichnet) aller Wörter, also die ungeordnete Menge aller Wörter, wobei Wörter – im Gegensatz zur herkömmlichen Menge – in dieser Multimenge auch mehrfach vorkommen. Wir berechnen also für jedes in einem der beiden Texte enthaltenen Wörter die Differenzen der Häufigkeiten $\text{freq}^T(w)$ für die Texte $T=GS$ (Gold-Standard) und $T=ET$ (erkannter Text), summieren diese auf, und normalisieren diese, in dem wir die Summe durch Gesamtanzahl aller Wörter im erkannten Text teilen:

$$\text{BER} := (\#_w \#_W |\text{freq}^{GS}(w) - \text{freq}^{ET}(w)|) / (\#_w \#_W \text{freq}^{ET}(w))$$

Frühere Analysen haben gezeigt, dass wir den Text als geeignet für die Analyse erachten können, wenn für eine Graphic Novel die BER kleiner als 40 ist. Auf

den erkannten Texten betrachten wir die Textähnlichkeit basierend auf einer euklidischen Vektordistanz der Dokument-Vektoren der Term-Dokument-Matrix, die jeweils die für jedes Dokument D die relative Vorkommenshäufigkeit $\text{tf}(D,t)$ der 2000 häufigsten Wörter t enthält. Bzgl. der visuellen Maße betrachten wir die mittlere Helligkeit jeder Seite, die Entropie und die Anzahl der Flächen als Maß für die visuelle Unruhe eines Bildes, sowie den Color Layout Descriptor und den Edge Histogram Descriptor des Standards MPEG7 (Martínez, Koenen & Pereira). Diese haben sich in früheren Arbeiten als vielversprechende Maße herausgestellt (Dunst & Hartel 2018a). Wann immer eine Dimensionsreduktion notwendig ist, führen wir diese mithilfe einer PCA durch. Um z.B. die textuellen und visuellen Maße kombiniert zu betrachten, haben wir – um einem Ungleichgewicht der 2000 Dimensionen für die 2000 häufigsten Wörter im Vergleich zu den ca. 40 visuellen Maßen entgegenzuwirken – zunächst via PCA die textuellen Dimensionen auf 40 reduziert. Anschließend haben wir die Dimensionen der textuellen PCA und die Dimensionen der visuellen Maße mit Hilfe einer weiteren PCA kombiniert. Zur Untersuchung signifikanter Zusammenhänge nutzen wir die ANOVA (ANalysis Of VAriance), die untersucht, ob die Varianz zwischen den Kategorien größer ist als die Varianz innerhalb der Kategorien.

Ergebnisse & Diskussion

Abbildung 2 stellt die Ergebnisse der multimodalen Stilometrie im Vergleich mit rein visuellen oder Textmaßen dar. Dabei zeigt sich, dass die Kombination von Bild- und Textanalyse nicht immer zum Erfolg führt. Zwar ergeben sich aus der Analyse beider Informationskanäle statistisch deutlichere Signifikanzen bei der Autor-Identifikation und Genreunterscheidung. Der Effekt ist allerdings gering. Im Fall der Klassifikation nach Originalsprache des Werks sowie unterschiedlicher Publikationsformen – etwa als Einzelband oder als fortgesetzte Serie – führt die Hinzunahme der Textanalyse derzeit nicht zu signifikanten Ergebnissen. Bei der Analyse unterschiedlicher Formen von Autorschaft (Einzelautor*innen, Zusammenarbeit von einer Autor*in und einer Illustrator*in und größeren Autor*innen-Teams) liegt das Resultat der Textanalyse weit über der Signifikanzgrenze. Die deutlich signifikanten Ergebnisse der visuellen Stilometrie setzen sich allerdings auch in der Kombination beider Kanäle durch. Insgesamt lässt sich sagen, dass trotz der gleichen Anzahl der verwendeten visuellen und Textmaße erstere derzeit aussagekräftiger erscheinen. Weiter erscheint es sinnvoll, die Analyse beider Informationskanäle immer auch einzeln zu betrachten und diese nicht immer sofort zu kombinieren.

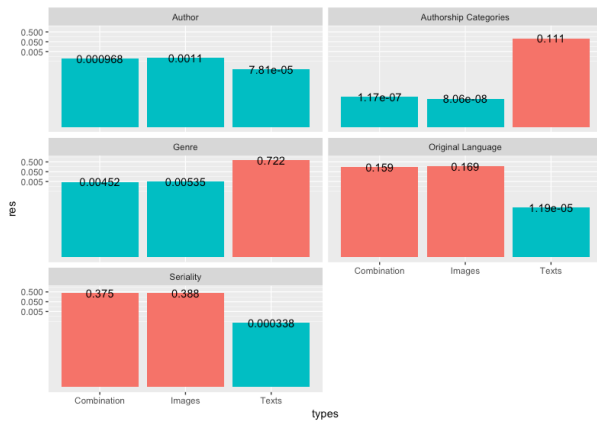


Abbildung 2. Zusammenfassung der stilometrischen Untersuchungen; Signifikanz ab $p < 0,05$

Wie bereits kurz erwähnt, zeigt sich für eine Klassifikation der entscheidende Einfluss der verwendeten Textmaße. Abbildung 3 stellt alle von uns untersuchten Werke in Streudiagrammen dar und unterscheiden diese farblich nach Originalsprache. Im Fall japanischer Manga handelt es sich hier außerdem um eine eigenständige Nationaltradition. Obwohl uns japanische und französischsprachige Werke in englischer Übersetzung vorliegen, führt nur die Textanalyse zu einer klaren Unterscheidung. Dies steht im klaren Gegensatz zu den Ergebnissen in Abbildung 1. Zwei potenzielle Ursachen können für dieses, auf den ersten Blick kontraintuitive, Ergebnis angeführt werden. Erstens erscheint es möglich, dass diese Unterscheidung eine Folge des Übersetzungsprozesses sind. Wahrscheinlicher ist, dass sich typische Merkmale der Texte von Manga auch in Übersetzung erhalten – in diesem Fall die Frequenz einzelner Wörter.

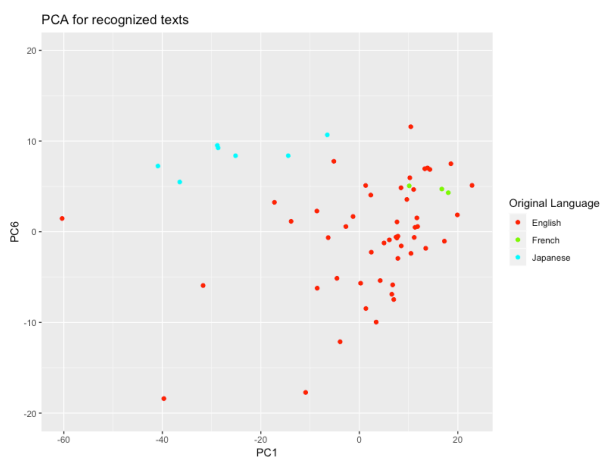


Abbildung 3. Streudiagramm basierend auf Textmaßen

Zusammenfassung & Ausblick

Wir haben erste Untersuchungen vorgestellt, die am Beispiel von Comicbüchern visuelle und Textmaße für eine multimodale Stilometrie kombinieren. Dabei handelt es sich, insbesondere in letzterem Fall, um sehr einfache Maße, die dem derzeitigen Stand der automatischen Texterkennung für Comicschriften geschuldet sind, und insgesamt um erste Pilotversuche. Wie sich zeigte, führt die Kombination der Text- und Bildebene in der Analyse bisher nicht immer zu besseren Ergebnissen. Allerdings ist dies trotz der geringen Anzahl der untersuchten Werke sowohl bei der Gattungsunterscheidung als auch bei der Autoridentifikation der Fall, für die in der Literaturwissenschaft seit längerem ähnliche Maße herangezogen werden. Insgesamt erscheint es sinnvoll, vor einer Kombination die Analyse der visuellen und textlichen Informationskanäle immer auch einzeln zu betrachten. Einen alternativen Zugang zu dem hier gewählten bietet die stilistische Klassifikation mit Hilfe neuronaler Netzwerke (für Comics: Laubrock & Dubray). Obwohl hier potenziell bessere Ergebnisse erzielt werden können, präferieren wir aus mehreren Gründen einen niederschweligen Ansatz: trotz der Zuhilfenahme der PCA in den hier abgebildeten Darstellungen versprechen wir uns von der Verwendung einzelner Maße eine bessere qualitative Interpretation. Zweitens ist dieser Zugang weniger datenhungrig und daher der geringen Anzahl an Werken in unserem Corpus angemessen. In einem nächsten Schritt wollen wir die Ergebnisse der Texterkennung verbessern. Dies wird es ermöglichen, zusätzliche Textmaße und Werke für unsere Analyse heranzuziehen und unsere Ergebnisse zu verbessern.

Bibliographie

- Baxter, Mike / Khitrova, Daria / Tsivian, Yuri (2016):** *A Numerate Film Theory? Cinematics looks at Griffith, Griffith Looks at Cinematics*, in: *Mise au Point* 8, <https://journals.openedition.org/map/2108> [letzter Zugriff 2. Januar 2019].
- Cutting, James / Brunick, Kaitlin / DeLong, Jordan / Iricinischi, Catalina / Candan, Ayse (2011):** *Quicker, faster, darker: Changes in Hollywood Film over 75 Years*, in: *i-Perception* 2: 569-576
- Dunst, A. / Hartel, R. (2018a):** *Automated Genre and Author Distinction in Comics*, DH 2018: Book of Abstracts 184-188.
- Dunst, Alexander / Hartel, Rita (2018b):** *The Quantitative Study of Comics: Towards a Visual Stylometry of Graphic Narrative*, in: **Dunst, Alexander / Laubrock, Jochen / Wildfeuer, Janina (Eds.): Empirical Comics Research: Digital, Cognitive, and Multimodal Methods**, New York: Routledge 43-61.
- Dunst, Alexander / Hartel, Rita / Laubrock, Jochen (2017):** *The Graphic Narrative Corpus (GNC): Design,*

Annotation, and Analysis for the Digital Humanities, in: Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition 15-20.

Flückiger, Barbar (2017): *Analysis of Film Colors in a Digital Humanities Perspective*, in: Frames 1, <http://framescinemajournal.com/article/analysis-of-film-colors-in-a-digital-humanities-perspective> [letzter Zugriff 2. Januar 2019].

Hartel, Rita / Dunst, Alexander (2019): *How Good is Good Enough? Establishing Quality Thresholds for the Automatic Text Analysis of Retro-Digitized Comics*, in: Proceedings of the Multimedia Modeling Conference (Springer Lecture Notes in Computer Science 11296), https://easychair.org/publications/preprint_open/Mdf2 [letzter Zugriff 2. Januar 2019].

Holmes, David (1998): *The Evolution of Stylometry in Humanities Scholarship*, in: Literary and Linguistic Computing 13: 111-17.

Laubrock, Jochen / Dubray, David (2018): *Computational Analysis and Visual Stilometry of Comics using Convolutional Neural Networks*, in: DH 2018: Book of Abstracts 228-231.

Manovich, Lev / Douglas, Jeremy / Zepel, Tara (2011): *How to Compare One Million Images*, <http://manovich.net/index.php/projects/how-to-compare> [letzter Zugriff 2. Januar 2019].

Martínez, J. / Koenen, R. / Pereira, F. (2002): *MPEG-7: the generic multimedia content description standard, part 1*, in: IEEE Multimedia 9: 78-87.

Qi, Hanchao / Taeb, Armeen / Hughes, Shannon (2013): *Visual stylometry using background selection and wavelet-HMT-based Fisher information distances for attribution and dating of impressionist paintings*, in: Signal Processing 93: 541-53.

Schmidt, Ben (2014): *Search for Structures in the Simpsons and everywhere else*, <http://benschmidt.org/2014/09/11/simpsons-2> [letzter Zugriff 2. Januar 2019].