

Zur Modellierung von Unsicherheit: Machine Learning und begriffliche Vagheit am Beispiel der Novellen im 19. Jahrhundert

Schröter, Julian

j.schroeter@lmu.de

LMU München, Deutschland

Einleitung

Die Modellierung von Vagheit und Unsicherheit in der sprachlichen Kategorisierung ist ein offenes Problem im Bereich des *Machine Learning*, eine drängende Herausforderung für die Computerlinguistik und eine Chance für die Rekonstruktion semiotisch komplexer kultureller Phänomene. Literaturgeschichte ist ein besonders interessanter Fall eines semiotisch komplexen kulturellen Systems. Pionierarbeit in der algorithmischen Modellierung kategorialer Vagheit in Fragen auf die Geschichte der Literatur wurde von Ted Underwood (2019b) als *Perspectival Modeling* geleistet, und zwar insbesondere an Beispielen historischer Genres. Underwoods wegweisender Gedanke besteht darin, aus der Not, dass der kulturelle Bias in den menschlich annotierten Trainingsdaten grundsätzlich zu einer Art *Human Bias* überwachter maschineller Lernverfahren führt, eine Tugend zu machen, indem Modelle auf den historischen Gebrauch von Gattungsbegriffen trainiert und vergleichend aufeinander beziehbar gemacht werden. Die Rekonstruktion von kulturellem Wissen über literarische Gattungen sollte sich daher als eine Rekonstruktion begrifflicher Unsicherheit verstehen. Modellierung auf der Grundlage überwachter maschineller Lernens im Feld der Digital Humanities ist, wie bereits Underwood (2019a) herausgestellt hat, besonders geeignet, Formen der Unsicherheit zu beschreiben. Seine Modellierungen haben jedoch die Ebene der Semantik selbst nicht berücksichtigt. Der vorliegende Beitrag möchte einen neuen Schritt in Richtung einer Modellierung kulturellen Wissens durch eine originelle Verbindung von drei Techniken oder Ideen gehen. Dafür gilt es in Abschnitt 2 Vagheit und begriffliche Offenheit auf der Ebene historischer Semantik zu modellieren. Diese Modellierung wird in Abschnitt 3 für eine neuartige Projektionsmethode in der konkreten historischen Analyse der Erzählprosa des 19. Jahrhunderts weiterentwickelt. Durchgeführt werden die entwickelten Methoden mit realen Daten aus einem neu aufgebauten Korpus von Jour-

nalprosa des 19. Jahrhunderts (Schröter et al., 2023). Abschnitt 4 führt mit einem weiteren Verfahren zur Messung komplexer Datenstrukturen eine plausibilisierende Triangulation durch, die im abschließenden Ausblick zu Fragen der Evaluation und Interpretierbarkeit der Ergebnisse führt. Die hierfür aufgegriffenen Techniken und Ideen sind die der perspektivischen Modellierung, also die Verwendung eines Unentscheidbarkeit berücksichtigenden Maßes für Vorhersagegenauigkeit im Bereich überwachter maschineller Lernens, und die Verwendung von Label-Distributionen anstelle binärer *Ground Truth* Label.

Modellierung von begrifflicher Vagheit als Prototypikalität

Die Sprachphilosophie und die literarische Gattungstheorie stellen mit der Prototypentheorie (Rosch, 1978; Taylor, 2007; Hempfer, 2010) und dem Familienähnlichkeitsbegriff nach Wittgensteins *Philosophischen Untersuchungen* (2001, Abs. 65–71), (Strube, 1986; Fishelov, 1991) zwei Modelle zur Verfügung, die es erlauben, begriffliche Vagheit abzubilden. Obwohl beide Modelle in Gattungsstudien der CLS aufgegriffen wurden (Calvo Tello, 2021; Henny-Kramer, 2023), wurde die gezielte Modellierung des in den Modellen konzipierten Vagheitscharakters noch nicht mit hinreichender Komplexität abgebildet. Auch wenn sich die Prototypentheorie bei Rosch (1978) aus Wittgensteins Überlegungen zur Familienähnlichkeit hergeleitet hat, kann man die beiden Konzeptualisierungen von Vagheit gut unterscheiden: Prototypikalität lässt sich als eine Form begrifflicher Vagheit begreifen, die dadurch zustande kommt, dass einzelne Instanzen (hier literarische Werkmanifestationen) mehr oder weniger nah am begrifflichen Zentrum eines Gattungsbegriffs sind. Einige Instanzen können dabei so weit vom begrifflichen Zentrum entfernt sein, dass sie konkurrierenden Kategorien (oder Gattungsbegriffen) ebenso nah oder sogar noch näher stehen als der Kategorie, der sie nominell zugerechnet werden. Die Prototypentheorie macht also das Szenario beschreibbar, dass ein Werk, das üblicherweise der Gattung der Novelle zugerechnet wird, dem begrifflichen Zentrum konkurrierender Gattungen wie dem Roman oder dem Märchen näher steht als der Novelle. Das Familienähnlichkeitsmodell, wie es in der Gattungstheorie etwa von Fishelov und Hempfer aufgegriffen wurde, zielt dagegen auf begriffliche Strukturen mit mehreren begrifflichen Zentren oder Kernen. Die Vagheit eines Gattungsbegriffs wie der Novelle im Sinn von Familienähnlichkeit kommt, wie Strubes (1986) Analyse zeigt, dadurch zustande, dass es mehrere Subtypen von Novellen gibt, wobei es – analog zu Wittgensteins Beispiel der völlig unterschiedlichen Arten von Spielen – möglich ist, dass einzelne Subtypen der Gattung derart unterschiedlich sind, dass es zwei Novellen geben kann, die keinerlei oder zumindest nur sehr wenige vermeintlich gattungsspezifische Merkmale gemeinsam haben. Die im Folgenden vorgestellte Mo-

dellierung widmet sich ausschließlich der strukturell durch Prototypikalität zustande kommenden Vagheit.

Die folgende Modellierung ist durch drei wesentliche Schritte gekennzeichnet: (1a) der Aufbau eines geeigneten Verfahrens maschinellen Lernens, das (1b) eine nutzbare Kennzahl für eine Bemessung der relativen Nähe und Distanz zum begrifflichen Zentrum bereitstellt, (2) die Erfassung eines Bereichs relativer Vagheit, und (3) ein Maß zur Bemessung des relativen Vagheitscharakters eines Gattungsbegriffs wie der Novelle. Zu (1): Gewählt wird (1a) Logistische Regression als klassisches Modell überwachten maschinellen Lernens. Auch wenn logistische Regression ein bereits in die Jahre gekommenes Lernverfahren darstellt, hat es zwei entscheidende Vorteile. Erstens die bereits von Underwood (2019a und 2019b) betonte gute Interpretierbarkeit der Features, und zweites die relative Vorhersagewahrscheinlichkeit für die Zuordnung einer Instanz zu einer der in Frage kommenden Klassen. Die grundlegende Idee zu Schritt (1b) ist es, relative Nähe einer Instanz zum begrifflichen Zentrum im Sinn von Prototypikalität durch die Vorhersagewahrscheinlichkeit für eine bestimmte Klasse auszudrücken.

Der entscheidende und neue Schritt ist nun (2) die Einführung eines Bereichs relativer Vagheit. Das vorgestellte Modell verwendet den sogenannten *C@I-Score* (Peñas und Rodrigo, 2011), eine Variante des *Accuracy score*, der Unentscheidbarkeit in einer mathematisch plausiblen Weise berücksichtigt. Seine Stärke liegt in der doppelten Funktion, begriffliche Unschärfe in der Weise interpretierbar zu machen, wie sie in der Literaturtheorie und Semantik seit den 1960er Jahren konzipiert wurde. Der *C@I-score* führt zusätzlich zur Anzahl der korrekten Prognosen k T und zur Gesamtzahl n aller Prognosen die Anzahl der Enthaltungen k U ein und hat die folgende Gleichung Gl. 1 (ebd., 1416).

$$c@1 = \frac{1}{n} \left(k_t + \frac{k_u}{2} + k_c \right) \quad (\text{Gl. 1 } c@1\text{-accuracy score, Peñas und Rodrigo 2011})$$

$$\sigma(x) = \frac{e^x}{1 + e^x} \quad (\text{Gl. 2 Logistische Funktion})$$

$$\log \left(\frac{p(x)}{1 - p(x)} \right) = \beta + w \cdot x \quad (\text{Gl. 3 Logistische Regression mit Link-Funktion})$$

$$\text{Verbesserungsrate} = \frac{c@1_{\text{trainiert}}}{\text{accuracy}_{\text{standard}}} \quad (\text{Gl. 4 Verbesserungsrate})$$

Im nächsten Schritt bedarf es einer Entscheidungsfunktion für Enthaltung. Hierfür wird mit logistischer Regression ein Algorithmus verwendet, der in der sigmoiden Funktion (Logistische Funktion, Gl. 2) mit der Link-Funktion (Gl. 3) über die Berechnung von Vorhersagewahrscheinlichkeiten läuft. Abbildung 1 zeigt, wie der Bereich einer Enthaltung symmetrisch um eine Vorhersagegenauigkeit von 0,5 gewählt werden kann.¹

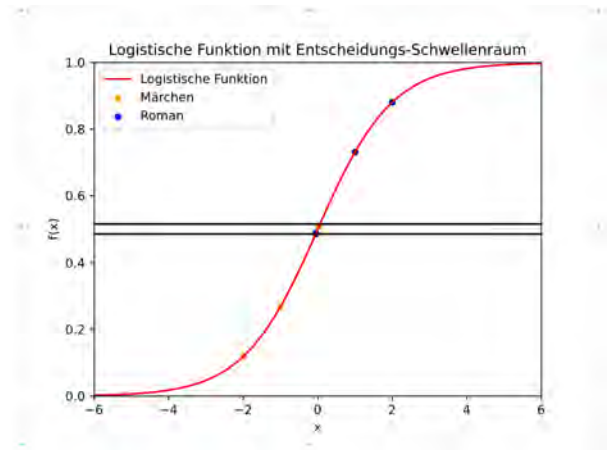


Abbildung 1: Logistische Funktion mit Schwellenraum der Enthaltung für simulierte Daten zur Klassifikation von Romanen versus Märchen.

Anstelle einer Entscheidungslinie bei einem Wert von $y = 0,5$ für die Vorhersagewahrscheinlichkeit wird eine Art Grauzone eingeführt, so dass bei Funktionswerten innerhalb dieses Bereichs anstelle einer binären Vorhersage für eine der beiden möglichen Klassen eine Enthaltung erfolgt. Die Enthaltung gibt lediglich eine epistemische Enthaltung des Modells wieder, ohne Implikation für den Status der Objekte. Was der *C@I-score* leistet, lässt sich über drei mathematische Eigenschaften ausdrücken: (1) Falls der Algorithmus jeden Fall entscheidet, so ist der *C@I-score* identisch mit dem Standard- *accuracy-score*: Diese Eigenschaft ist wichtig für die hier gewählte Anwendung, denn nur wenn diese Eigenschaft erfüllt ist, lassen sich beide Varianten aufeinander beziehen. Der *Accuracy score* ist der besondere Fall des *C@I-scores*, bei dem jeder Fall binär entschieden wird. (2) Eine korrekte Entscheidung führt immer zu einem besseren Wert als eine Enthaltung. (3) Eine Enthaltung führt zu einem besseren Wert als eine falsche Entscheidung.

Die optimale Breite des Enthaltungsbereichs kann in Form einer Gridsuche ermittelt werden. Abbildung 2 zeigt den Graphen, der diese Gridsuche am Beispiel eines Modells zur Klassifikation von Romanen vs. Märchen als Funktion abbildet, bei der für das trainierte Modell und die verfügbaren Validierungsdaten jeweils zu einem sukzessive ausgeweiteten Enthaltungsbereich für Vorhersagewahrscheinlichkeiten zwischen 0 und 1 der zugehörige *C@I-score* als Funktionswert bestimmt wird.²

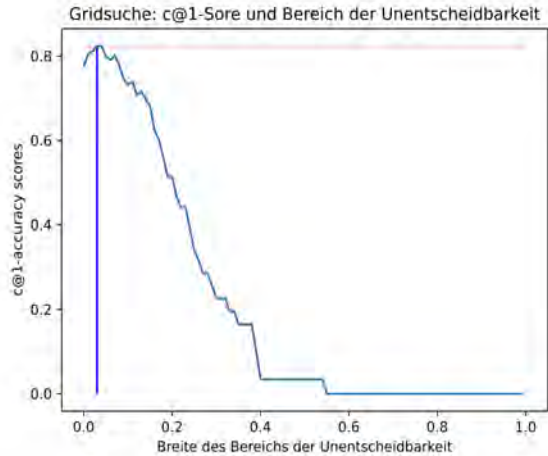


Abbildung 2: Bestimmung eines optimalen Bereichs der Unterscheidbarkeit für die Entscheidungsfunktion mit einer Gridsuche für die Klassifikation von Märchen versus Romanen, hier aber auf einem Feature-Set der 2500 häufigsten Wörter im Korpus.

Zu (3): Zur Berechnung des Grads an begrifflicher Vagheit, die auf diese Weise algorithmisch modelliert wird, stehen mehrere Maße als potenzielle Kandidaten zur Verfügung: (a) die Breite des optimalen Enthaltungsbereichs und (b) die *Verbesserungsrate* (s. Gl. 4) die mit einem $C@1$ -Score bei optimalem Enthaltungsbereich ($C@1$ optimal) gegenüber einem klassischen *Accuracy score* (*accuracy standard*) erzielt werden kann. Voraussetzung für die Verwendung des *Accuracy scores* ein hierbei ein Sampling, das je gleich große Validierungsstichproben für die zu vergleichenden Gruppen zieht.

In einem Prozess interner Evaluation wurden Daten und simuliert, um das Verhalten der Berechnung optimaler Enthaltungsbreiten besser zu verstehen. Die Ergebnisse werden in den Visualisierungen aus Abbildung 3 dargestellt. Hierfür wurde Vagheit im Sinn einer Kombination von Familienähnlichkeit und potenzieller Prototypikalität dadurch simuliert, dass in einigen Fällen jede Klasse nur ein Feature-Cluster mit mehreren informativen Features und mehr oder weniger großer Varianz sowie Störrauschen (im Sinn einer Annäherung an Prototypikalität) erzeugt wurde (linke Spalte) und in einigen Fällen jede Klasse mit mehreren Feature-Clustern (2 Cluster je Klasse in der mittleren Spalte sowie 5 Cluster je Klasse in der rechten Spalte). Jedes Simulation wurde 1000-fach iteriert. Die Boxplots in der Mitte zeigen die Verteilung der optimalen Breite des optimalen Enthaltungsbereichs über alle Iterationen. Die untere Zeile zeigt die Verteilung der Verbesserungsrate. Es zeichnet sich ab, dass die Verbesserungsrate die robustere und aussagekräftigere Kennzahl ist, um komplexe Merkmalsbündel mit komplizierten Mehrfachclustern innerhalb der Klasse abzubilden. Wenn man annimmt, dass Kategorien, die durch mehrere interne Cluster gegliedert sind und dadurch vage werden, dem Familienähnlichkeitsbegriff entsprechen, dann lässt sich auf einer strukturellen Ebene sehen, dass die Verbesserungsrate anhand des $C@1$ -scores bei einem optimalen Enthaltungsbereich eine Modellierung

von Prototypikalität darstellt, die Vagheit im Sinn der Familienähnlichkeit einzufangen erlaubt. Der Vergleich von Simulation und realen Daten erlaubt Schlussfolgerungen für künftige Studien. Obwohl perspektivische Modellierung auf Klassifikationsaufgaben beruht, die über die Logik von »ja oder nein« und nicht über die von »mehr oder weniger« laufen, ist das »Mehr oder Weniger« in einige Klassifikationsalgorithmen sozusagen eingebaut. Der $C@1$ -score ist für sich genommen noch nicht aussagekräftig. Insbesondere direkte Vergleiche beispielsweise des $C@1$ -score s mit in anderen Szenarien ermittelten klassischen *Accuracy score* s sind nicht sinnvoll. Die nützlichste Kennzahl ist die der Verbesserungsrate.

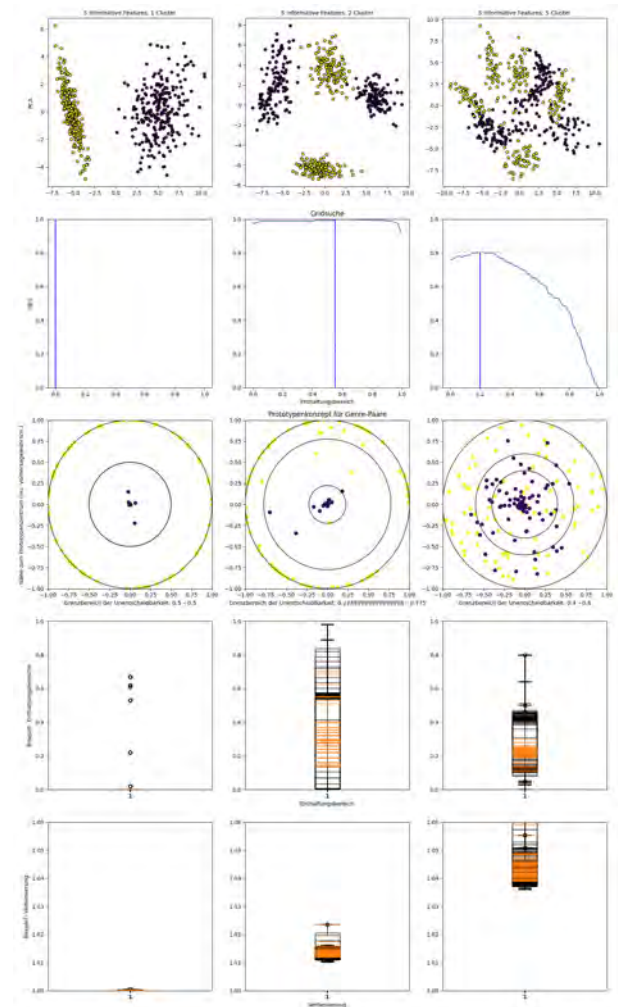


Abbildung 3: Simulation mit Default-Einstellung: 10 Features, $class_sep=1$, $flyp_y=0,0001$, Anordnung auf einem Hyperwürfel mit der Funktion *make_classification* in *Scikit-Learn*. Die kreisförmige Projektion zeigt die um ein Prototypenzentrum angeordneten Datenpunkte mit den eingezeichneten Grenzen für den optimalen Enthaltungsbereich als Kreislinien, als Projektionsmethode vorgestellt in Schröter (2022).

Projektionen von Prototypikalität und Fallstudien zur Novelle

Mit dem entwickelten Verfahren lassen sich aufschlussreiche Anomalien historischer Gattungen sichtbar machen, die wichtige, aber im kodifizierten »Literaturwissen« nach wie vor marginalisierte Positionen in der historischen Gattungsforschung stützen (Meyer, 1987; Lukas, 1998). Hierfür wurde eine Projektionsmethode entwickelt, die jeweils eine Gattung im Verhältnis zu den beiden Nachbargattungen visualisiert. Abbildung 4 zeigt die Prototypikalität nach paarweise konstruierten Gattungsunterscheidungen. Die Achsen zeigen jeweils die invertierten Vorhersagewahrscheinlichkeiten (siehe y-Achse aus Abbildung 1). Dargestellt werden die Texte einer Gattung aus dem Set der *out-of-sample*-Validierungsdaten mit den über Bootstrapping berechneten Vorhersagewahrscheinlichkeiten mit einem 1000-fach iterierten Klassifikationstask. Jede der Visualisierungen zeigt ausschließlich die Texte einer Gattung, und zwar derjenigen, deren Zentrum sich jeweils unten links befindet. Die Enthaltungsbereiche zu jeweiligen Nachbargattungen sind im Farbcode der jeweiligen Komplementärgattung (grün für Erzählungen und rot für Novellen) eingetragen. Die Projektion macht sichtbar, dass Erzählungen (rechts) über das gesamte 19. Jahrhundert hinweg betrachtet – ein homogenes Abgrenzungsverhalten zu ihren beiden jeweiligen Nachbargattungen haben. Novellen dagegen sind in ihrem Abgrenzungsverhalten anomal: Als Novellen publizierte Texte (links), die das Modell im Verhältnis zu Erzählungen eindeutig als Novellen erkennt, sind typischerweise in der Unterscheidung zu Romanen nicht novellen- sondern romanartig, und umgekehrt.

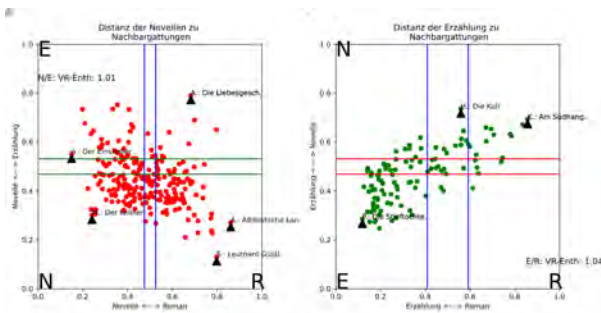


Abbildung 4: Perspektivische Modellierung von Prototypikalität: Novellen (N) (links) und Erzählungen (E) (rechts) in Abgrenzung zu je zwei Kontrastgattungen (E und R bzw. N und R) unter Verwendung der invertierten Vorhersagewahrscheinlichkeiten (1-p) auf beiden Achsen; mit Angabe der Verbesserungsrate durch Enthaltung (VR-Enth.) und mit optimalen Enthaltungsbereichen. Siehe die Skripte »plot_perspectival_[Gattungen]_triad[Zeitraum].py«.

Triangulationen

Die im Detail in Spezialfragen historischer Gattungsforschung führenden Ergebnisse lassen sich methodologisch

auf gewinnbringende Weise für Fragen der algorithmischen Modellierung von Vagheit innerhalb historischer Semantiken generalisieren, wenn es gelingt, diese Modellierung von Unbestimmtheit so zu plausibilisieren, dass die sichtbar gemachte Vagheit tatsächlich einem Ausmaß erwartbarer Vagheit in der historischen Semantik entspricht. Bis hierher lässt sich feststellen, dass das entwickelte Modell so etwas wie die Struktur der Prototypikalität im Raum algorithmischer Konzeptualisierung abbildet. Ob die Resultate indes dem entsprechen, was man in der Linguistik sowie in der Literaturwissenschaft unter Prototypen einer Gattung oder Kategorie sowie unter dem Grad an prototypikalischer begrifflicher Struktur eines bestimmten Gattungsbegriffs versteht, bedarf der Kombination (oder Triangulation) mit weiterer philologischer, evtl. auch psychologischer Evidenz. Daher wird eine Möglichkeit der Triangulation in Form eines Vergleichs der entwickelten und auf dem Konzept der Unentscheidbarkeit beruhenden Verbesserungsrate mit einer alternativen Konzeptualisierung von Vagheit diskutiert: Zumindest für die simulierten Fälle sollten die durch mehrere Feature-Cluster (s. Abb. 3 Mitte und Rechts) erzeugten Klassen mit potenzieller kategorialer Vagheit in Form von ‚bluriness‘ durch nicht-lineare Algorithmen überwachten maschinellen Lernens (wie Neuronale Netze, Random Forests und nicht-lineare SVM) besser klassifizierbar sein als lineare Modelle (wie logistische Regression oder eine SVM mit linearem Kernel). Tabelle 1 zeigt, dass sich die Verbesserungsrate durch Unentscheidbarkeit mit C@1-score und die Verbesserungsrate durch nicht-lineare Modellierung sogar für die realen Daten, also die Erzählprosa des 19. Jahrhunderts und deren Gattungszuordnungen, proportional zueinander verhalten. Dieses Ergebnis stärkt die Annahme, dass die hier vorgestellte Modellierung von begrifflicher Vagheit geeignet ist, ein Strukturverhältnis zwischen Feature-Verteilung und Klassenzugehörigkeit zu erfassen.

Tabelle 1: Klassifikation mit Vergleich von nicht-linearem (Random Forest) und linearem Modell (logistische Regression) für Gattungspaare mit (nach Z-Transformation) skalierten Daten (mit: 2500 lemmatisierten *mfw*). In Klammern sind jeweils die Werte angegeben, die auf einer Berechnung nach Z-Transformation (Standard-Skalierung der Daten) beruhen. Siehe die Skripte »run_classification_rnd_forest_model-comp-ttest_[Gattungspaar].py«.

Gattungspaar	Verbesserung durch nicht-lineare Modelle	Verbesserung durch C@1-score	
Roman	Erzählung	1,16 (1,14)	1,039
Novelle	Roman	1,12 (1,21)	1,025
Novelle	Erzählung	1,00 (1,04)	1,014

Ausblick: Evaluation und *Human Label Variation*

Zuletzt lassen sich zwei weitere wichtige Schritte für künftige Anschlussforschung präsentieren und so zur Diskussion stellen: Zum einen die Möglichkeit und vor allem die Herausforderungen einer Evaluation anhand psycholinguistischer Evidenz, indem menschliche Urteile zu mehr oder weniger prototypischen Beispielen von Kate-

goriezugehörigkeit herangezogen werden. Zum anderen ist der Umgang mit uneindeutigen Kategorisierungen ein offenes Problem im Arbeitsfeld des *Machine Learning*, an dem im Moment auch im Bereich der Entwicklung von LLMs im Bereich des NLP intensiv geforscht wird.³ Der Vortrag skizziert deshalb abschließend einen Vorschlag im Anschluss an Peterson et al. (2019, 9619), wonach anstelle von *Ground Truth Labels* eine Verteilung von Labels für das Training Neuronaler Netze so verwendet werden kann, dass die relative Vagheit und Unsicherheit bereits in der Praxis der historisch uneinheitlichen Gattungszuordnung repräsentiert ist. Im Zuge einer davon unabhängigen Validierung anhand von Daten, die das Modell beim Training nicht gesehen hat, kann mit entsprechend komplexer Modellierung geprüft werden, ob das Modell die relative Vagheit und auch die der historischen Gattungszuordnung reproduzieren kann. Ziel des Beitrags ist eine Modellierung der theoretischen Struktur von begrifflicher Vagheit im Sinn der Prototypensemantik, die zu einem besseren Verständnis von semantischer Vagheit in historischen Kontexten beitragen kann. Der vorgestellte Ansatz möchte zu einer Lösung einer zentralen Herausforderung im Feld des *Machine Learning* beitragen, nämlich, mit Unsicherheit der Kategorisierung rational umzugehen.

Fußnoten

1. Alle Methoden und folgenden Visualisierungen können mit den Python-Skripten, Funktionen, Klassen und Methoden in der Bibliothek PyNovellaHistory(Schröter, 2024) nachvollzogen werden. Die folgenden Verweise auf relevante Python-Skripte sind mit Links hinterlegt.
2. Die Visualisierung dieser Gridsuche kann in den Skripten mit der Benennungsregel »run_classi_fic_principle_d_sampling_[Gattungspaar].py« repliziert werden. Der optimale *C@1-score* von 0,82 wird in diesem Beispiel bei einer Breite des Enthaltungsbereichs von 0,03 erreicht. Die bis hier vorgestellte Modellierung wurde, der grundlegenden Idee nach, bei einer Poster-Präsentation (Schröter, 2022) in Urbana-Champaign (Illinois) vorgestellt. Alle folgenden Arbeitsschritte sind noch nicht öffentlich gemachte Weiterentwicklungen.
3. So etwa im Rahmen des ERC-Projets DIALECT, <https://dialect-erc.github.io/>.

Bibliographie

- Calvo Tello, José.** 2021. *The Novel in the Spanish Silver Age. A Digital Analysis of Genre Using Machine Learning*. Bielefeld: Bielefeld University Press.
- Fishelov, David.** 1991. „Genre Theory and Family Resemblance – Revisited“. *Poetics* 20: 123–38.
- Hempfer, Klaus W.** 2010. „Zum begrifflichen Status der Gattungsbegriffe: Von ›Klassen‹ zu ›Familienähnlichkeiten‹ und ›Prototypen‹“. *Zeitschrift für französische Sprache und Literatur* 120: 14–32.
- Henny-Kramer, Ulrike.** 2023. *Genre Analysis and Corpus Design: Nineteenth-Century Spanish-American Novels (1830–1910)*. Schriften Des Instituts Für Dokumentologie Und Editorik, 17. Berlin: IDE.
- Lukas, Wolfgang.** 1998. „Novellistik“. In *Zwischen Restauration und Revolution 1815–1848*, hg. von Gert Sautermeister und Ulrich Schmid. Hansers Sozialgeschichte der deutschen Literatur 5. München: Hanser, 251–280.
- Meyer, Reinhart.** 1987. *Novelle und Journal, I: Titel und Normen: Untersuchungen zur Terminologie der Journalprosa, zu ihren Tendenzen, Verhältnissen und Bedingungen*. Stuttgart: Steiner.
- Peñas, Anselmo, und Alvaro Rodrigo.** 2011. „A Simple Measure to Assess Non-response“. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 1415–1424.
- Peterson, Joshua, Ruairidh Battleday, Thomas Griffiths, und Olga Russakovsky.** 2019. „Human Uncertainty Makes Classification More Robust“. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 9616–9625.
- Rosch, Eleanor.** 1978. „Principles of Categorization“. In *Cognition and Categorization*, herausgegeben von Eleanor Rosch und Barbara B. Lloyd. Hillsdale (NJ): Erlbaum, 27–48.
- Schröter, Julian.** 2022. „Modeling Prototypicality for Genre Concepts (Poster presentation at the iSchool Showcase, University of Illinois at Urbana-Champaign)“. Zenodo. <https://doi.org/10.5281/zenodo.7945007>.
- Schröter, Julian .** 2024. „PyNovellaHistory: Python Code for the Project on the history of the German 19th- century novella“. Zenodo. <https://doi.org/10.5281/zenodo.7945348>.
- Schröter, Julian, Johannes Leitgeb, und Theresa Valta.** 2023. „Ein digitales Korpus der Novellen und Journalprosa des 19. Jahrhunderts: Herausforderungen der Metadatenerschließung“. *Dariah Working Papers* 46. <https://doi.org/10.47952/gro-publ-131>.
- Strube, Werner.** 1986. „Sprachanalytisch-philosophische Typologie literaturwissenschaftlicher Begriffe“. In *Zur Terminologie der Literaturwissenschaft*, hg. von Christian Wagenknecht. Stuttgart: Metzler, 35–49.
- Taylor, John R.** 2007. *Linguistic categorization*. 3. ed., repr. Oxford textbooks in linguistics. Oxford: Oxford Univ. Press.
- Underwood, Ted.** 2019a. „Algorithmic Modeling“. In *The Shape of Data in Digital Humanities: Modeling Texts and Text Based Resources*, hg. v. Fotis Jannidis und Julia Flanders. London: Routledge, 250–263.
- Underwood, Ted .** 2019b. *Distant Horizons. Digital Evidence and Literary Change*. Chicago, London: The University of Chicago Press.
- Wittgenstein, Ludwig.** 2001. *Philosophische Untersuchungen*. Frankfurt am Main: Suhrkamp.