

JANINA JACKE (KIEL)

# Die (computationelle?) Operationalisierung unzuverlässigen Erzählens

Ein Beitrag zur Theorie und Methodik literaturwissenschaftlichen  
Interpretierens

Ziel des Projekts *Computer-aided Analysis of Unreliability and Truth in Fiction* (kurz: CAUTION) ist die Operationalisierung des erzähltheoretischen Konzepts *unzuverlässiges Erzählen*. Beim unzuverlässigen Erzählen handelt es sich um ein Konzept, das ein narratives Phänomen mit einer komplexen und heterogenen Extension fassen soll – grob gesprochen, dient es der Kategorisierung von Erzählinstanzen in fiktionalen Texten, deren Äußerungen nicht für bare Münze zu nehmen sind. CAUTION befasst sich dabei (zunächst) nur mit einem Teilbereich des unzuverlässigen Erzählens, nämlich mit faktenbezogener beziehungsweise mimetischer Unzuverlässigkeit, bei der eine Diskrepanz zwischen den Behauptungen der Erzählinstanz und den Fakten der fiktiven Welt vorliegt, beziehungsweise – noch genauer – mit dem speziellen Fall der Vergabe *inkorrekturer Informationen*. Unter Operationalisierung ist hier die Identifikation der Handlungsschritte zu verstehen, die notwendig sind, um festzustellen, ob das durch ein literaturwissenschaftliches Konzept beschriebene literarische Phänomen in einem Text vorliegt oder nicht.

In diesem Rahmen sollen zuvorderst zwei (miteinander verbundene, aber unterscheidbare) offene theoretische Fragen im Hinblick auf unzuverlässiges Erzählen beantwortet werden. Die erste Frage lautet: Unter welchen Bedingungen und inwiefern handelt es sich bei der Feststellung unzuverlässigen Erzählens um einen interpretativen Vorgang? Diese Frage knüpft daran an, dass unzuverlässiges Erzählen oft als grundsätzlich interpretationsabhängiges Phänomen bezeichnet wird – wobei in der Regel allerdings weder überzeugende theoretische Gründe angegeben noch empirische Studien durchgeführt werden, um diese Annahme zu belegen. Die zweite Frage lautet: Welche Rolle spielen textuelle Indikatoren bei der Feststellung unzuverlässigen Erzählens? Hintergrund ist hier die Tatsache, dass in der Unzuverlässigkeitsforschung häufig unterschiedliche Textmerkmale genannt und in Listen zusammengestellt werden, die auf unzuverlässiges Erzählen hindeuten können, darunter beispielsweise linguistische, strukturelle oder inhaltliche Phänomene. Diese Indikatoren werden meist pauschal – ganz in Übereinstimmung mit der angenommenen grundsätzlichen Interpretationsabhängigkeit – als weder notwendig noch hinreichend für unzuverlässiges Erzählen eingeordnet, wiederum ohne eine angemessen differenzierte theoretische Begründung oder empirische Überprüfung.

Die Operationalisierung unzuverlässigen Erzählens erfolgt in CAUTION im Rahmen eines experimentellen methodischen Settings, bei dem verschiedenartige computergestützte Zugänge verfolgt werden, deren Ergebnisse miteinander in Beziehung gesetzt und deren Möglichkeiten und Grenzen jeweils reflektiert werden. Das Projekt ist dementsprechend interdisziplinär angelegt und wird gemeinsam von einem literaturwissenschaftlichen (PI Janina Jacke, Kiel) und einem computerlinguistischen Projektpartner (PI Jonas Kuhn, Stuttgart) durchgeführt.

Basierend auf dieser allgemeinen Projektkonzeption besteht die Hoffnung, dass CAUTION – über die konkret auf das unzuverlässige Erzählen bezogenen Fragestellungen hinaus – auch Bausteine zur Beantwortung zweier weiterer verknüpfter, aber unterscheidbarer Fragen mit größerer Reichweite beitragen kann. Die erste dieser Fragen ist von allgemein-literaturtheoretischem Interesse und lautet: Welche Rolle spielen deskriptiv feststellbare Textmerkmale bei der Interpretation literarischer Texte? Die zweite Frage ist methodologischer Natur und ist insbesondere für das Feld der Digital Humanities relevant: Wie können computergestützte Methoden fruchtbar eingesetzt werden, um die Beantwortung interpretationsabhängiger Fragen von geisteswissenschaftlichem Interesse zu unterstützen?

Im Folgenden sollen die in CAUTION zum Einsatz kommenden computergestützten Zugänge zur Operationalisierung des Konzepts ›unzuverlässiges Erzählen‹ kurz vorgestellt und ihr Zusammenhang und ihre Relevanz begründet werden. Die in CAUTION zum Einsatz kommenden Zugänge sind: (1) automatische Annotation ausgewählter sprachlicher Indikatoren, (2) intuitive manuelle Annotation (a) ausgewählter inhaltlicher Indikatoren und (b) inkorrektur Äußerungen, (3) kollaborative inhaltspezifische Interpretation unter Nutzung von Argumentationsdokumentation und -visualisierung, (4) automatische Textklassifikation mithilfe von Deep Learning.

(1) Automatische Annotation ausgewählter sprachlicher Indikatoren: In den Feldern der Computerlinguistik und der computationellen Literaturwissenschaft existieren bereits Methoden zur automatischen Feststellung solcher (oder ähnlicher) sprachlicher Phänomene, wie sie auch als Indikatoren für unzuverlässiges Erzählen genannt werden. Beispiele sind expressive Sprache sowie ein gehäuftes Auftreten von Modaladverbien und direkter Adressat\*innenansprache. Es bietet sich deswegen an, entsprechende computationelle Modelle zusammenzustellen, ihre Anwendbarkeit auf literarische Texte zu testen und zu evaluieren, ob sie tatsächlich geeignet sind, die sprachlichen Phänomene zu fassen, die als Unzuverlässigkeitsindikatoren gehandelt werden. Ist dies der Fall, werden die Modelle auf ein kleines literarisches Korpus angewandt, das die zentrale Textgrundlage für die ersten drei Zugänge bildet. Dabei handelt es sich um neun deutschsprachige fiktionale Erzählungen kurzer bis mittlerer Länge aus dem Zeitraum zwischen 1800 und heute. Das Korpus enthält Erzählungen, die in der literaturwissenschaftlichen Forschung eher als unzuverlässig gehandelt werden, und solche, bei denen das eher nicht der Fall ist. Durch einen Abgleich dieser automatischen Annotationsdaten mit den in den übrigen Ansätzen generierten Daten lassen sich erste Tendenzen hinsichtlich der Frage erkennen, ob die sprachlichen Indikatoren tatsächlich in einem relevanten Maße mit (festgestellter) Unzuverlässigkeit beziehungsweise weiteren verwandten Phänomenen korrelieren.

(2) Intuitive manuelle Annotation: In der Forschung zu unzuverlässigem Erzählen wird meist nicht genauer spezifiziert, wie die angenommene Beziehung zwischen (sprachlichen) Indikatoren und Unzuverlässigkeit beschaffen ist. Eine Analyse ergibt, dass dabei meist ein gedanklicher Zwischenschritt gemacht wird: Bestimmte sprachliche Merkmale werden als Hinweis auf eine bestimmte Eigenschaft der Erzählinstanz (z. B.

Charaktermerkmale, mentale Zustände oder Intentionen) verstanden, welche dann wiederum einen Grund dafür darstellen kann, dass die Erzählinstanz inkorrekte Äußerungen über die fiktive Welt tätigt, also unzuverlässig ist. So scheint beispielsweise ein vermehrtes Auftreten von einschränkenden epistemischen Modaladverbien auf eine unsichere Erzählinstanz hinzudeuten, der möglicherweise mit höherer Wahrscheinlichkeit Fehler unterlaufen. Eine interessante Frage scheint nun zu sein, wie die tatsächliche Relation zwischen sprachlichen Phänomenen und den relevanten Erzähler\*inneneigenschaften einerseits und diesen Eigenschaften und erzählerischer Unzuverlässigkeit andererseits beschaffen ist. Zu diesem Zweck werden (a) einige relevante Erzähler\*inneneigenschaften in den Korpus-texten annotiert. Da die Identifikation dieser Eigenschaften bereits Interpretation erfordert, bietet sich kein automatischer Zugang an. Stattdessen werden die Eigenschaften – satzbasiert und nach möglichst intuitivem Leseverstehen – von Literaturwissenschaftler\*innen annotiert, wobei jeder Text von mindestens zwei Annotator\*innen bearbeitet wird. Im selben methodischen Rahmen werden in den Korpus-texten zudem (b) diejenigen Sätze annotiert, in denen die Erzählinstanz inkorrekte Äußerungen über die fiktive Welt tätigt. Die Annotator\*innen haben im Rahmen beider Annotationsaufgaben die Möglichkeit, neben klaren Zuordnungen auch deutlich zu machen, an welchen Stellen sie sich hinsichtlich der Einordnung unsicher sind. Zudem können Begründungen für Annotationsentscheidungen notiert werden.

Die Evaluation der so generierten Daten bietet nun bereits einiges Potenzial zur Beantwortung der zentralen Fragen: Zum einen lassen sich erste Erkenntnisse darüber erzielen, in welcher Beziehung deskriptiv feststellbare sprachliche Indikatoren, potenziell relevante Erzähler\*inneneigenschaften und zugeschriebene Unzuverlässigkeit zueinander stehen. Zum anderen kann evaluiert werden, wie stark und an welchen Stellen sich manuelle Annotationen unterscheiden, was wiederum Implikationen für die Interpretationsabhängigkeit der annotierten Phänomene zu haben scheint.

(3) Kollaborative inhaltspezifisierende Interpretation unter Nutzung von Argumentationsdokumentation und -visualisierung: Wenn einem Phänomen Interpretationsabhängigkeit zugeschrieben wird, dann ist darunter in der Regel nicht zuvorderst der Umstand zu verstehen, dass *de facto* unterschiedliche Meinungen darüber existieren, ob es in einem Text vorliegt oder nicht, sondern dass unterschiedliche *legitime* Interpretationen möglich sind, die zu unterschiedlichen Diagnosen hinsichtlich seines Vorliegens führen. Soll geprüft werden, ob die Feststellung von unzuverlässigem Erzählen in diesem Sinne interpretationsabhängig ist, sind intuitive Annotationen nicht ausreichend. Deswegen werden die vormaligen Annotator\*innen dazu angehalten, für die Korpus-texte jeweils inhaltspezifisierende Interpretationen zu entwickeln, also eine Hypothese darüber aufzustellen, was in der fiktiven Welt eines Textes der Fall ist. Denn unzuverlässiges Erzählen ergibt sich aus dem Widerspruch zwischen (oft deskriptiv feststellbarer) Erzähler\*innenäußerung und (interpretativ rekonstruierbarer) fiktiver Welt. Diese Hypothese soll dann unter Berücksichtigung konkurrierender Optionen begründet und verteidigt werden. Um diese komplexe Aufgabe übersichtlicher und die Ergebnisse vergleichbar zu gestalten, folgt sie einem vorgegebenen Schema: Zunächst werden die relevanten offenen Fragen hinsichtlich der fiktiven Welt eines Textes und die jeweils möglich erscheinenden Antwortoptionen zusammengestellt. Im Anschluss wird eine möglichst kohärente inhaltspezifisierende Interpretation entwickelt, indem für jede Frage eine Antwort ausgewählt wird. Für jede Antwort werden Argumente angeführt – dabei werden die Relationen zwischen Thesen und stützenden Annahmen mithilfe von Argumentbäumen im Rahmen von Argumentationsvisualisierungssoftware verdeutlicht. Die vormaligen Annotator\*innen vergleichen und diskutieren ihre Interpretationen und

Argumentationen und versuchen im Falle abweichender Einschätzungen, die anderen argumentativ zu überzeugen. Dabei besteht das Ziel allerdings keineswegs darin, Einigkeit zu erzwingen. Auf der Basis der hinreichend diskutierten Interpretationen werden dann die vormals intuitiven Annotationen inkorrektur Äußerungen der Erzählinstanz ein zweites Mal ausgeführt. Die so generierten Daten bieten nun zum einen die Möglichkeit herauszufinden, wie groß die Unterschiede zwischen den Annotator\*innen hinsichtlich der Feststellung unzuverlässigen Erzählens nach umfassender Reflexion noch sind. Zum anderen können durch Analyse der Argumentbäume diejenigen Aspekte identifiziert werden, die für persistente Interpretationsunterschiede verantwortlich sind, und es lässt sich feststellen, inwieweit für die Begründung der relevanten Interpretationshypothesen auf sprachliche Texteigenschaften und Erzähler\*inneneigenschaften zurückgegriffen wird.

(4) Automatische Textklassifikation mithilfe von Deep Learning: Ein letzter, experimenteller (und literaturtheoretisch am wenigsten informierter) Zugang besteht darin, größere Korpora von kanonisch als unzuverlässig bzw. zuverlässig erzählten Texten zusammenzustellen und mithilfe so genannter neuronaler Netzwerke computationell analysieren zu lassen. Die Idee ist dabei, dass der Computer auf subtile und zugleich komplexe Texteneigenschaften anspringt, die dem menschlichen Auge womöglich entgehen, und es auf diese Weise möglich ist, nicht-vorklassifizierte Texte automatisch als unzuverlässig beziehungsweise zuverlässig erzählt einzuordnen. Um auch im Rahmen dieses Ansatzes nach Möglichkeit literaturtheoretisch relevante Erkenntnisse zu erzielen, ist es wichtig, dass mit Verfahren erklärbarer künstlicher Intelligenz (*explainable AI*) gearbeitet wird, die einen Einblick in Prozesse und Methoden erlauben.

In ihrer Kombination ermöglichen die in CAUTION eingesetzten Verfahren nicht nur interessante theoretische und methodologische Erkenntnisse hinsichtlich der eingangs genannten vier Fragen. Darüber hinaus zeichnet sich ab, dass aus dem Projekt computationelle Modelle zumindest zur automatischen Erkennung von im Zusammenhang mit unzuverlässigem Erzählen relevanten Komponenten hervorgehen werden, die für die Exploration und Analyse größerer Textkorpora eingesetzt werden können. Zu beachten ist in beiden Zusammenhängen allerdings, dass die Verallgemeinerbarkeit der Ergebnisse aufgrund der kleinen Korpusgröße noch stark eingeschränkt ist.