



**MODELL ROMANTiK**  
Variation · Reichweite · Aktualität

Udo Hahn

**Impuls zu Ted Underwoods**  
*Distant Horizons. Digital Evidence and Literary Change*

**Publikation**

Erstpublikation als *Wissenschaftlicher Impuls* auf der Plattform „Gestern | Romantik | Heute“  
Vorlage: Udo Hahn

URL: <https://www.gestern-romantik-heute.uni-jena.de/wissenschaft/artikel/impuls-zu-ted-underwoods-distant-horizons-digital-evidence-and-literary-change>

Veröffentlicht zuerst am 05. April 2022

Herausgegeben von Sandra Kerschbaumer, Romy Langeheine und Alexander Pappe

DOI: 10.22032/dbt.61459

Lizenz:



**Empfohlene Zitierweise**

Beim Zitieren empfehlen wir nach der URL-Angabe das Datum Ihres letzten Besuchs dieser Online-Adresse sowie die DOI anzugeben:

Udo Hahn: Impuls zu Ted Underwoods *Distant Horizons. Digital Evidence and Literary Change*. In: *Gestern | Romantik | Heute*. <https://www.gestern-romantik-heute.uni-jena.de/wissenschaft/artikel/impuls-zu-ted-underwoods-distant-horizons-digital-evidence-and-literary-change> (Datum Ihres letzten Besuchs). <https://doi.org/10.22032/dbt.61459>.

**Impuls zu Ted Underwoods**  
*Distant Horizons. Digital Evidence and Literary Change*

Udo Hahn

Underwoods *Distant Horizons* ist ein Buch über *distant reading*, eine Form der literarischen Studienkultur, die sich – initiiert durch Franco Morettis Pamphlet *Conjectures on World Literature* (2000) – mittlerweile vor allem als empirische, quantitativ fundierte Literaturwissenschaft charakterisieren lässt, die mit großen Mengen von Textdaten operiert. Sie steht damit in einem spannungsgeladenen Zielkonflikt zum weithin dominierenden *close reading*, das auf der sorgfältigen subjektiven Interpretation ausgewählter Passagen literarischer Werke beruht und eher auf autor- bzw. werktypische Singularitäten und Partikularitäten als verallgemeinernde Analysen zielt.

Underwoods Buch steht exemplarisch für die Korrektur von verbreiteten Fehlannahmen über das empirisch-quantitative Forschungsprogramm, das oft mit von Computern durchgeführten einfachen Zähloperationen gleichgesetzt wird: Wie häufig treten etwa (Farb-)Adjektive, Personalpronomen oder Konjunktive in den Werken eines Autors auf? Diese rein deskriptive Sicht greift jedoch viel zu kurz, denn Computer werden im weiteren wissenschaftlichen Kontext weniger zum reinen Zählen von festgelegten Merkmalen eingesetzt (das auch), sondern in viel größerem Ausmaß mittlerweile als Hypothesentest-Technologie verstanden, also zur Prüfung von gemeinhin komplexen Aussagen, Modellen und Theorien unter Verwendung großer Datenbestände. Diese sind auch für die Geisteswissenschaften mittlerweile in digitalen Bibliotheken (wie *Hathi Trust* (<https://www.hathitrust.org/>) oder *Europeana* (<https://www.europeana.eu/de>)) in Form von viele Millionen bibliothekarische Einheiten umfassenden Kollektionen verfügbar.

*Distant Horizons* gliedert sich in fünf annähernd gleichgewichtige Kapitel. Das erste führt in die von Underwood vertretene Methodenkonzeption ein, die nächsten drei wenden diese auf konkrete literaturwissenschaftliche Problemstellungen an (Genre-Unterscheidungen, Identifizierung von literarischem Prestige und die Rolle des Geschlechts in Büchern und bei Autoren), das fünfte reflektiert die zugrundeliegende und an den oben genannten drei Beispielen exemplifizierte quantitative Methodenkonzeption. Die folgenden zwei Appendices könnten im Grunde Kapitel 6 und 7 sein, da sie Underwoods Perspektive auf Daten und Methoden (erneut) thematisieren und kritisch beleuchten.

Kapitel 1 setzt auf dem Selbstverständnis der Literaturwissenschaft (wie es Underwood wahrnimmt) auf – der Stabilität literarischer, auch zeitlich klar demarkierter Epochen wie „English Romanticism“ (ca. 1800–1840) oder Klassifikationen wie „Viktorianische Novellen“ sowie den damit jeweils verbundenen Werkekanon. Underwood wirft die Frage auf: Verstehen wir überhaupt die grundlegenden Konturen und *langfristigen* Trends der Literaturgeschichte jenseits zeitlicher und epochenspezifischer Kartographierung? Er bezweifelt dies unter Hinweis auf globale Muster entlang wesentlich längerer, periodenüberspannender Zeitlinien, für die die Literaturwissenschaft derzeit noch nicht einmal Benennungen parat hat. Ihre Aufdeckung soll nicht als Ersatz für bekannte, eher lokale Kategorisierungen oder interpretationsleitende Periodenschemata dienen, sondern als deren Ergänzung und so eine Einordnung in völlig neue, umfassendere Beschreibungsdimensionen ermöglichen.

Wie kommt man diesen langläufigen Mustern näher? Underwoods erster Vorschlag lautet: **Messen!** Etwa die Häufigkeit von Farbbezeichnungen in ca. 350 Werken zwischen 1700–1920, die eine signifikante, stetig steigende Zunahme seit 1750 in diesen Werken erfährt – genauer: durchschnittlich eine Verdreifachung vom 18. ins frühe 20. Jahrhundert. Eine solche punktuelle Aussage dürfte jedoch klassisch denkende Literaturwissenschaftler wenig elektrisieren und ist auch nur der erste Schritt in seiner *distant reading*-Programmatik.

Underwoods nächster Vorschlag lautet daher: **Korrelieren!** Welche statistischen Effekte treten neben der Zunahme von Farbbezeichnungen in der Literatur des 19. und 20. Jahrhunderts ebenfalls auf? Etwa die Zunahme jedweder Form physischer Beschreibungen (Benennungen von Körperteilen, Handlungs- und Wahrnehmungsverben, konkrete Adjektive, etwa zur Farbenbezeichnung, sogenannte *hard seeds*) verbunden mit der gleichzeitigen Abnahme der Frequenz von Bezeichnungen abstrakter Werte (Schönheit, Trauer, Liebe usw.) – dies spiegelt den allgemeinen Trend *from telling to showing*, von direkt ausgedrückten sozialen Einschätzungen zu solchen, die indirekt über physische Details evoziert werden, wider. Dass diese *hard seeds* auch zur statistisch verlässlichen Genre-Unterscheidung fiktionaler von nicht-fiktionaler Literatur (hier werden exemplarisch Biographien betrachtet) taugen, wird in der Folge weiter vertieft (die Daten gründen in mehr als 3.800 Werken, die in der *Hathi Trust*-Digitalbibliothek vorliegen). Evidenzen dieser Art beruhen auf Wortlisten, die durch semantische Kategorien (wie „abstrakt“ vs. „konkret“) weiter strukturiert sind, aber doch stets unvollständig und an ihren kategorialen Grenzen schwer zu begründen sind. Wie aber sollen diese quantitativen Bewegungen, die gerade skizziert wurden, literaturkritisch plausibel interpretiert werden (Hinweis: statistische Korrelationen sind kein Beleg für Kausalität)?

Underwoods dritter Vorschlag zielt auf die Umkehrung der bisherigen Vorgehensweise: **Modellieren! Hypothesen testen!** Anstatt mit Messungen zu beginnen, um dann Korrelationsmuster zu entdecken und schließlich zu fragen, was sie bedeuten, wird die Reihenfolge der einzelnen Schritte umgedreht. Die Bedeutung eines Musters wird als Hypothese (als Instanz eines Modells) formuliert, über Messungen der in der Hypothese auftretenden Kategorien (Merkmale in Form von Variablen; bei Underwood sind dies überwiegend einzelne Lexeme, deren Vorkommenshäufigkeit bestimmt wird) getestet und so auf Gültigkeit geprüft. Am Beispiel eines statistischen Modells für die Genre-Differenzierung zwischen fiktionaler und nicht-fiktionaler Literatur (Biographien) werden Genre-Prädiktoren (Variablen) über Wahrscheinlichkeiten verknüpft, also Relationen zwischen verschiedenen Variablen statistisch bestimmt. Dieses Vorgehen ersetzt die Betrachtung einzelner Messachsen (wie die singuläre Häufigkeit von Farbbezeichnungen). Woher kommt aber die *ground truth*, anhand derer solche Modelle auf ihre Gültigkeit evaluiert werden können?

Underwoods Antwort darauf ist (nur auf den ersten Blick) verblüffend einfach: etwa von Bibliothekaren, die Bücher nach den Kategorien „fiktional“ vs. „non-fiktional“ klassifizieren. Oder von entsprechenden Einordnungen von Literaturwissenschaftlern in Leselisten für ihre Seminare oder in Einführungsbüchern. Diese Auswahlen definieren einen Werkekanon als direkte Reflektion des gebündelten, überwiegend konsensualen Vorwissens von Genreexperten. Die Aufgabe eines statistischen Modells ist dann, diese Vorklassifikationen – mit welchen Merkmalsgruppen auch immer – zu reproduzieren und damit daraus abgeleitete statistische Zusammenhänge im Vorwissen dieser Experten, der *scientific community*, zu verankern. Wenn diese Maßgabe erfüllt ist, sind statistische Aussagen über *hard seeds* ein empirisches Korrelat für eine Interpretation statt nur ein deskriptives Faktum, dessen Interpretation offen ist. Da statistische Modelle Stärkegrade (ein Kontinuum) von Kategorisierungen ausdrücken können, eignen sie sich zudem für die Beschreibung der Unschärferänder entsprechender diskreter Kategorien, die (nicht nur) literaturwissenschaftlichen Kategorien eigen sind. Für den Aufbau

des statistischen Modells verwendet der Autor (regularisierte) logistische Regression als etablierte Methode des maschinellen Lernens. Sie ist im Stande, diejenigen Wörter auszuwählen, die am stärksten beide Genres diskriminieren, sodass zwar Wortlisten manuell bereitgestellt werden müssen, die Relevanz jedes einzelnen Wortes – sein Beitrag zur Genre-Diskriminierung – aber automatisch (algorithmisch) berechnet wird.

Underwood weist mit Recht darauf hin, dass das einfache Messen von Worthäufigkeiten kognitiv einsichtiger und damit einfacher vermittelbar sei als die Beschreibung eines statistischen Modells (und seiner technischen Voraussetzungen in Form der logistischen Regression, Regularisierung, Kreuzklassifikation oder dem Modell-Tuning, die er allesamt nur kurz streift). Aber für die Einfachheit der Häufigkeitszählung zahlt man den Preis der offenen Interpretierbarkeit. Dies ist bei statistischen Modellen anders, da die Interpretation als Ausgangspunkt fixiert wird und man diese Interpretation an empirischen Evidenzen festzumachen versucht – beispielsweise durch Worthäufigkeitszählungen, die für stilometrische Experimente trotz ihrer stupenden Einfachheit in der Tat eine verblüffend große diskriminative Kraft entfalten. Auch technisch ungeübte Leser können den roten Faden dieser methodologischen Prolegomena gut nachvollziehen, die immer eng an der literaturwissenschaftlichen Motivation Underwoods, Grundlagen für eine drei Jahrhunderte umfassende Literaturhistorie zu legen, geführt werden.

Mit dieser Programmatik widmet sich Underwood in Kapitel 2 der Lebensspanne von Genres, einem in der Literaturkritik äußerst schillernden Begriff. Underwood schlägt vor, beispielsweise Jules Vernes Romane gegen die seiner Zeitgenossen des 19. Jahrhunderts zu vergleichen (sie sollten strukturell voneinander abweichen) und diesen Vergleich dann auf *Science Fiction*-Literatur des 20. Jahrhunderts auszuweiten (sie sollten strukturell ähnlich zu Vernes Arbeiten sein). Gelingt dieser Vergleich, ist das Genre „*Science Fiction*“ auf der Grundlage der gemessenen Ähnlichkeit und in sie einfließenden Variablen(stärken) strukturell charakterisiert und somit seine Existenz über zwei Jahrhunderte mittels prädiktiver statistischer Modelle nachgewiesen. Diese Form der „*perspektivischen Modellierung*“, auf kontextbezogenem Vergleich basierend, ist das Kernelement von Underwoods Arbeitsweise – für Genres, (und später) für literarisches Prestige und die Rolle des Geschlechts in der Literatur.

Immer wieder betont der Autor, dass die hier entwickelten Modelle (nur) auf textuellen Evidenzen beruhen – Erwartungen der Leser, Praktiken der Literaturproduktion u. ä. bleiben außen vor, was deren (potenzielle) Relevanz jedoch nicht in Abrede stellt. Diese scheinbare Schwäche ist aber eher eine Stärke des quantitativ-digitalen Ansatzes und letztlich jeden empirisch-experimentellen Vorgehens, denn sie lässt anderen Forschern die Gelegenheit, diese bislang nicht untersuchten Faktoren in breiter ausgelegte prädiktive Modelle zu integrieren und diese dann zu testen. Prädiktive Modelle stellen dabei probabilistische Beziehungen zwischen Variablen her, die keine noch so bemühte Definition (etwa von „Genre“) erfassen kann. Ein Hinweis darauf, dass begriffslastige Wissenschaften ein immanentes Erkenntnisproblem und entsprechende Definitionsdebatten begrenzten Nutzen haben.

Die Bezugspunkte für sein empirisches Vorgehen sind wie schon erwähnt autoritative Listen von Forschern, Katalogzuordnungen von Bibliothekaren, Klassifikationen von Rezensenten in (Literatur-)Zeitschriften usw., d. h., sie reflektieren im Kern konsensuale soziale und ästhetische Realitäten (in Form von hohen Überlappungsraten) der Literaturgruppierung. Die Auswahl der Merkmale (Underwood fokussiert auf einfache, ausschließlich lexikalische Merkmale wie Worthäufigkeit, durchschnittliche Wortlänge) hat dabei weit weniger Einfluss auf Genre-Unterscheidungen als gemeinhin gedacht, da die Genre-Zugehörigkeit redundant über viele linguistische Ebenen verteilt ist – d. h., andere Merkmalswahlen führen zu (sehr) ähnlichen Ergebnissen. Solche Effekte treten an vielen Stellen in den Wissenschaften auf und werden

etwa unter dem Motto „*simplicity counts!*“ oder der Strategie des *Occam's razor* subsumiert. Machen Sie selbst einen Test, um die Robustheit dieser Annahme zu prüfen: Lesen Sie aus einem Buch zwei oder drei zufällig ausgewählte Paragraphen, und Sie wissen (beispielsweise ohne Kenntnis der komplexen Plot-Struktur) mit großer Sicherheit, ob Sie einen Detektiv- oder Science-Fiction-Roman in der Hand halten.

Im analytischen Teil von Kap. 2 werden dann drei Genres diskutiert: Detektivromane, *Science-Fiction* und *Gothic novels*. Die Form der subtilen Vergleiche innerhalb und zwischen diesen Genres liefert eindrucksvolle Daten für deren Stabilität über längere zeitliche Horizonte und ihre wechselseitige Unterscheidbarkeit. Gleichzeitig werden damit eine Fülle literaturwissenschaftlicher Annahmen zu historischen Entwicklungen dieser Genres und Epochenunterschiede infrage gestellt – etwa die auf eine Generation von Lesern beschränkte Kurzläufigkeit von Genres, deren zunehmende Konsolidierung im Laufe der Zeit oder die Disparität von kulturellen Formen in der Genreentwicklung. Dass insbesondere genrebezogene Kurzspannenvergleiche (etwa über 30 Jahre) genauso stabil wie Langzeitvergleiche (etwa über 150 bis 200 Jahre) sind, ist ein Ergebnis Underwoods, das viele Standardannahmen in der Literaturwissenschaft erschüttert. Diese Befunde gelten vor allem für Detektivromane und Science-Fiction-Literatur, die *Gothic novels* sind ein zu diffuses Konstrukt mit divergierenden Themen und Settings (ausgehend von *Dracula* zählen hierzu Geister- und Horrorgeschichten bis hin zu Inkarnationsmärchen, die oft mit dem Topos des Liebeswerbens verknüpft sind), die die Gültigkeit dieser Hauptaussagen Underwoods nicht tragen. Man könnte auch offensiver urteilen und sagen, dass im Unterschied zur klaren Konturierung von Detektiv- und Science-Fiction-Romanen die Literatur, die unter dem Etikett „*Gothic novels*“ zusammengeführt wird, wegen der durch die Modelle signalisierten großen Divergenzen eher keinen literarisch relevanten gemeinsamen Nenner hat und deren vereinheitlichende Zusammenführung unter einem Namen eher unplausibel erscheint. Da die Parameter und Merkmale für diese Modellierung offen dargelegt werden, ist diese Einschätzung (im Gegensatz zum geschlossenen subjektiven Interpretieren) intersubjektiv kritisierbar. Andere Parametrisierungen der Modelle oder die Hinzunahme neuer, bislang unbeachteter Merkmalsgruppen können also durchaus noch eine gemeinsame Kennlinie für *Gothic novels* herausarbeiten.

Wo Kapitel 2 Aspekte der textuellen Genre-Konstitution behandelt und damit die Gruppierung von Leserschaften von Texten, wendet sich Kapitel 3 weiteren textexternen, überwiegend aber sozialen Aspekten der Literatur zu; Faktoren, die man als literarisches Prestige umschreiben kann und somit mehr auf Rezensenten und Käufer als auf Autoren verweisen. Auch hier zählt der Autor empirische Evidenzen auf, dass sich gebräuchliche 20–30-jährige Kurzzeit-Epochen (wie Romantizismus, Viktorianischer Realismus, Ästhetizismus, Naturalismus oder Modernismus), die in der Aufeinanderfolge aus Aufmerksamkeitserheischungsgründen oft als „*revolutionär*“ etikettiert werden, besser in Langzeit-Beschreibungen im Kontext eines Jahrhunderts (erfasst wird der Zeitraum von 1820 bis 1950) mit noch höherem Erklärungswert für stabile literarische Bewertungsstandards durch eher *langsame* Änderungsraten beschreiben lassen (also ein hohes Maß an Persistenz bzw. Dauerhaftigkeit aufweisen). Beachtenswert ist hier, dass soziale Prominenz direkt mit literarischem Sprachgebrauch in Zusammenhang gebracht und nicht reflexartig auf Verkaufszahlen, Bestseller-Listen u. ä. rekurriert (auch diese Faktoren zieht Underwood übrigens in Betracht).

Angesichts vieler offensichtlicher Möglichkeiten, literarisches Prestige zu messen (Literaturpreise, Leselisten in Seminaren usw.), entscheidet sich Underwood für einen Modellierungsansatz, der Prestige am Gegenstand von Rezensionen in elitären Periodika (gemeint sind damit Zeitschriften wie *The Atlantic*, *The New Yorker*, *Fortnightly Review* usw.); er nennt dies den Nachweis literarischer Distinktion) festmacht – über wen gesprochen wird, der ist relevant. Die

Erhebung des Prestiges wird für zwei literarische Formen, Poesie und Fiktion, betrieben, und zwar wiederum auf der Grundlage von Worthäufigkeitszählungen (andere Kriterien wie Rhythmik und Versmaß für Poesie oder bedingte Entropie (aus der Signal- bzw. Informationstheorie entlehnt) für Fiktion liefern ganz im Sinne der schon oben angestellten reduktiven Überlegungen keinen signifikant steigenden Beitrag zur Genauigkeit der Ergebnisse). Zwar fällt die Erkennungsgenauigkeit für prestigehaltige Poesie (auf 80 Prozent) und für prestigehaltige Fiktion (gar auf 73 Prozent) – für die Genres wurden 93 Prozent (für Detektivromane) und 91 Prozent (für *Science-Fiction*) ermittelt, doch sollte der vorgestellte Ansatz angesichts der geringen Komplexität der zugrundeliegenden Modelle für nicht-triviale Entscheidungsprobleme (im Fall der Poesie: was gibt einem poetischen Text Prominenz und dem Autor literarisches Prestige?) als eine gelungene, aber eben nur erste Modellnäherung verstanden werden. Verzerrungen bei der Textauswahl (Stichprobe) scheinen hier, wie Experimente Hinweise geben, übrigens keine unbedeutende Rolle zu spielen. Wie „hart“ die hier beschriebene Problemstellung wirklich ist, illustriert Underwood an einem interessanten Nebenexperiment: Experten (Studenten mit einem ersten Abschluss in und Dozenten der Literaturwissenschaft) erhielten Ausschnitte von rezensierten (prestigehaltigen) und zufällig ausgewählten (vermutlich prestigelosen) Texten, um das gleiche Problem zu lösen (stammt ein Text aus der Menge rezensierter Texte?). Die Versuchspersonen erzielten nur eine Genauigkeit von 64 Prozent – und eine Reihe dieser Entscheidungen dürfte eher auf Wiedererkennung des individuellen Textes als auf allgemeinen literaturtheoretischen Erwägungen beruhen.

Einem naheliegenden Einwand, das Modell assoziiere literarisches Prestige lediglich mit der Wahrscheinlichkeit, in elitären Journalen rezensiert zu werden, und blende andere relevante Faktoren aus, tritt Underwood durch die Betrachtung der Variablen „Nationalität“ (ist „Britishness“ ein einflussreicher Faktor für Prestige in der angloamerikanischen Literaturszene?) und „Geschlecht“ (werden männliche Autoren bei der Zuteilung von Prestige bevorzugt?) entgegen. Anhand einfachster statistischer Normalisierung des Datensatzes (durch Rebalancierung) kann er zeigen, dass (mit der Ausnahme des Prestiges für britische Poesie, die einen nationalen Präferenzbias indizieren, aber auch Effekte des verzerrten Samplings der Datensätze widerspiegeln könnte) dies in Bezug auf beide Variablen allgemein nicht gilt – weder Nationalität (mit der gerade angesprochenen Ausnahme) noch Geschlechtszugehörigkeit ändern die Grundmuster, die das „Rezensionsmodell“ vorgibt. Ein interessanter Seitenaspekt ist auch der empirische Befund, dass literarisches Prestige und Markterfolg (gemessen an Bestseller-Listen und mehreren Auflagen eines Buches) sich mit Beginn des 20. Jahrhunderts entkoppeln.

Kapitel 4 beschäftigt sich mit der Rolle des Geschlechts in fiktionaler Literatur von 1780 bis heute und greift damit, anders als die vorangegangenen Kapitel, tief in das Innenleben der behandelten Werke hinein – Gegenstand der Betrachtung sind nun tausende fiktionale Charaktere. Die Daten wurden auf der Basis von knapp 94.000 Büchern aus der *Hathi Trust*-Bibliothek gewonnen, eine Größenordnung, die weit außerhalb von Menschen erfassbarer Literaturmengen liegt, aber für Computer problemlos zu prozessieren ist. Für die Erkennung von Geschlechtszuordnungen zu einzelnen erzählten Charakteren reicht eine einfache Häufigkeitszählung lexikalischer Elemente nicht mehr aus; Underwood verwendet zum Tracking von Charakteren in fiktionalen Werken, d. h. der Sammlung von Eigenschaften und Attributen, die einer Figur vom Autor (etwa über sie beschreibende Adjektive, von ihnen als Subjekt dominierte Verben usw.) zugewiesen werden, die Software BookNLP (für Interessierte: <https://github.com/dbamman/book-nlp>). Auf der Basis solcher Merkmalsmengen lernt BookNLP automatisch prototypische Repräsentationen von männlichen und weiblichen Figuren und ist in der Umkehrung dann auch in der Lage, in dem System bislang unbekanntem

Texten das Geschlecht einer Figur zu präzisieren, und zwar mit einer Präzision von über 90 Prozent.

Unter diesen Rahmenbedingungen ermittelt BookNLP zwei zentrale empirische Befunde. Die Fähigkeit, Männer von Frauen zu unterscheiden, nimmt ab der Mitte des 19. Jahrhunderts mit zunehmender Nähe zur Jetztzeit zum Teil dramatisch ab. Der zweite Trend zeigt, dass die schiere Anzahl und (bezogen auf Buchseiten) der Raum für die fiktionale Beschreibung von Frauen oder Mädchen als Charaktere im Vergleich zu Männern oder Jungen permanent abnimmt. Der Befund gilt übrigens auch dann, wenn männliche und weibliche Autoren (Letztere sind im Originaldatenset unterrepräsentiert) perfekt balanciert sind, wobei männliche Autoren klarere Geschlechterzuordnungen erlauben als weibliche. Weibliche Autoren weisen darüber hinaus ein besprochenes Geschlechterverhältnis zwischen Frauen und Männern von 1:1 auf, männliche Autoren dagegen eines von 1:3 (fokussieren also erheblich stärker auf Männer).

Ein interessanter Nebenschauplatz sind aus den Experimenten einfach ableitbare verbale Signale für Weiblichkeit/Männlichkeit, und zwar als Häufigkeit eines Lexems zur Charakterisierung von Frauen subtrahiert von seiner Häufigkeit, mit der es Männer charakterisiert, und der Darstellung dieser Differenz über die betrachteten Zeitachsen. Underwood diskutiert eine große Menge von empirischen Details, aus denen ich eine anekdotisch herausheben möchte: die Verben zur Heiterkeit, die die Volatilität von Geschlechterzuordnungen deutlich machen. Um 1950 etwa identifiziert „*smile*“ und „*laugh*“ in hohem Maße Frauen, während Männer in ebenso hohem Maße durch „*chuckle*“ und mehr noch durch „*grin*“ denotiert werden – diese Geschlechterdifferenz ist hier übersignifikant, mündet aber zum Ende des 20. Jahrhunderts hin wieder in eine Konvergenz. Analoges Material findet sich zur Beschreibung menschlicher Physis und Kleidung („*eyes*“, „*hair*“ (das aktuell stärkste Weiblichkeitssignal); „*chest*“, „*pocket*“ (ein sehr starkes, wie „*chest*“, Männlichkeitssignal)), das Befunde aus Kapitel 1 zur Charakterisierung fiktionaler Literatur über physische Merkmale weiter stützt.

Kapitel 5 ist einigen Implikationen des von Underwood propagierten Ansatzes gewidmet und mit „*The Risks of Distant Reading*“ überschrieben. Es behandelt zunächst die Voraussetzungen (technisches Know-how, etwa im Bereich Statistik und maschinelles Lernen) und Konsequenzen einer Digitalen Geisteswissenschaft im Kontrast zur traditionell betriebenen. Hier werden zunächst Verdrängungsängste thematisiert, *distant reading* etwa als Dambruch charakterisiert, dessen nachfolgende Flut die hermeneutisch ausgerichtete, subjektiv geprägte *close reading*-Community aus den Geisteswissenschaften spülen könnte. Diese Position teilt Underwood in keiner Weise: Er betont die großen interpretatorischen Leistungen des *close reading*, sieht die Quantifizierung eher als Ergänzung denn als Substitution. Eine wissenschaftliche Revolution (im Sinne von Thomas Kuhn) will Underwood (im Gegensatz zu Moretti) nicht anzetteln, dafür argumentiert er in seinem ganzen Buch viel zu ausgleichend, ja vorsichtig, aber seine Nadelstiche gegen den literaturwissenschaftlichen *common sense* setzt er mit unaufgeregtem Habitus, professioneller Kompetenz und Belesenheit sowie letztlich klarer Verankerung in seinen Datensätzen und Modellen, die Neues zu sagen haben.

Im Grunde strebt er aber eine Versöhnung beider Welten an, der quantitativen und der qualitativen, abgeleitet aus einem für ihn fundamentalen Anliegen der Literaturkritik. Dem chirurgischen Messer von analytischen Zahlen hält Underwood die Pflicht der Literaturkritik entgegen, die Freude am Lesen von Poesie, Novellen und Schauspielen zu erhalten (und Hinweise auf Gründe für das Vergnügen daran zu geben). Das etablierte Format wissenschaftlicher Texte aus dem Dunstkreis der empirischen Wissenschaften (Methode – Experiment – Ergebnisse – Schlussfolgerungen) ist ihm für seine Wissenschaft ein Gräuelformat und sein Buch

entspricht dem durch seinen diskursiven, an vielen Stellen essayistischen Stil. Letztlich hält Underwood ein Plädoyer für die Koexistenz von *close* und *distant reading* als Methoden der Gewinnung literaturwissenschaftlicher und literaturkritischer Erkenntnisse. Und das erste große Risiko des *distant reading* besteht für Underwood darin, sich in der Menge von kleinteiligen empirisch-quantitativen Befunden zu verlieren, statt sie einsichtig zu bündeln und (im Sinne der von ihm geforderten „*pleasurable resonance*“) mit dem Lesevergnügen des Publikums an der Literatur in Einklang zu bringen.

Das zweite Risiko sieht er in der Vermengung, ja Gleichsetzung, von neuen (quantitativen) interpretativen Methoden mit den neuen digitalen Technologien. Er macht deutlich, dass für ihn *distant reading* eine neue Form der Interpretation darstellt, weniger eine Technologie zur Berechnung von Interpretationen. Zum Schluss fasst Underwood sein Paradigma des *distant reading* als historische Wissenschaft mit drei zentralen Begriffen zusammen: historische Untersuchungen können als **Experiment** angelegt sein, in denen Hypothesen an vorab festgelegten Stichproben getestet werden. Diese **Stichproben** aus der historischen Vergangenheit dienen dazu, (statistische) **Modelle**, die (probabilistische) Beziehungen zwischen Begriffen (Variablen) beschreiben, einer systematischen Prüfung zu unterziehen.

Mit diesem methodologischen Credo thematisiert Underwood ein drittes Risiko: Wie können diese Anforderungen seriös in die curricularen Strukturen und institutionellen Organisationsformen von Universitäten, hier insbesondere literaturwissenschaftliche Institute, integriert werden? Die erforderlichen inhaltlichen Schwerpunkte sind unzweideutig: tiefes Verständnis der Literaturgeschichte (mindestens) einer nationalen Kultur, Grundlagen der Statistik und des maschinellen Lernens, Praxis des Programmierens, Kenntnisse aus den Bereichen Linguistik und Sozialwissenschaften. Mit einem gewissen Erschrecken diskutiert Underwood eine Alternativlösung: Können Studenten der Informatik bzw. Informationswissenschaft, deren Curriculum mehrere dieser Inhalte bereits abdeckt, mit einer Einführung in die Literaturwissenschaft *distant reading* nicht ebenso bearbeiten? Um die Dramatik dieser Überlegungen noch mehr zuzuspitzen, weist Underwood darauf hin, dass die Literaturwissenschaft kein akademisches Erstzugriffsrecht auf die literarische Analytik der menschlichen Vergangenheit hat. Schottet sie sich von erkenntnistiftenden methodischen Innovationen ab, läuft sie Gefahr, in mittelfristiger Perspektive marginalisiert zu werden. Oder von anderen Strömungen, wie die *Digital Humanities* (denen Underwood jedoch eine zu technologische Perspektive und multidisziplinäre Diffusheit unterstellt), gar substituiert zu werden. Für Institutionen, die sich den „*uncountable particulars*“ (Levine) verschrieben haben, kämen curriculare Spuren von Mathematik und Informatik tatsächlich einer inhaltlichen Revolution gleich, deren Eintrittswahrscheinlichkeit der Autor aber als eher gering ansieht.

Die Arbeiten von Underwood sind daten- und softwaregetrieben; beides ist in einem frei zugänglichen Repositorium bei *Zenodo* (<https://zenodo.org/record/1206318>) niedergelegt (und wer der Programmiersprache Python kundig ist, kann auch ausführlich mit dem Programmcode, der den hier erörterten Ergebnissen zugrunde liegt, arbeiten). Wer alternative Quellen zu der von Underwood vertretenen Programmatik sucht, mag einen Blick auf die Webseiten des *Chicago Text Lab* (<https://lucian.uchicago.edu/blogs/literarynetworks/>) oder des *Stanford Literary Lab* (<https://litlab.stanford.edu/>) werfen.

Dieses Buch beschließen zwei Appendices, die mehr eine Fortsetzung der inhaltlichen Argumentation sind als technische Handreichungen. In Appendix A zu Daten diskutiert der Autor die Rolle, die seine jeweiligen Datensätze (und ihre Fehler und Beschränkungen) spielen, und wie die Art der Fragestellung deren Neuzusammenstellung (Rekombination) bestimmt hat. In Appendix B zu Methoden werden zunächst traditionelle Fehlannahmen zur Rolle von Zahlen

und zu quantitativ gestütztem Schlussfolgern auf Seiten der Geisteswissenschaft diskutiert (Determinismus, Objektivität usw.), um dann auf die besondere Komplexität literaturwissenschaftlich relevanter statistischer Modelle zu verweisen, die oft tausende von Variablen enthalten können – Dimensionen, die nur noch durch die Benutzung von leistungsstarken Rechnern bearbeitet und interpretatorisch kontrolliert werden können. Dimensionen, die aber auch Probleme für die statistische Theorie hinter dem maschinellen Lernen schufen (der sog. *bias-variance trade-off*) und erst gelöst werden mussten (etwa das Problem der dünnen Datenbelegung einzelner Variablen, d. h., liegen mehr Variablen als einzelne Datenpunkte vor, führt zum *over-fitting* und damit fehlender Generalisierungsleistung der Modelle). Eine wichtige wissenschaftstheoretische Konsequenz dieser neuen algorithmischen Lerntheorie ist, dass entsprechende Modelle zwar zur Prädiktion taugen (wie im Buch an vielen Stellen gezeigt), aber kaum mehr zur Erklärung – zumindest in Extremfällen größter Modellkomplexität. Diese (scheinbare) Einschränkung passt aber perfekt in Underwoods Programm – und dies ist die Prädiktion historisch langer Prozesse.

Underwood legt seinen Studien vergleichsweise einfache statistische Verfahren und Methoden des maschinellen Lernens zugrunde (logistische Regression) sowie lexikalische Klassifikationssysteme wie die des *General Inquirer* (1968), die einer mittlerweile veralteten, allerersten Generation von automatischen Inhaltsanalysesystemen entstammen (aktueller wären heute etwa digitale Lexika wie *WordNet* oder *Wiktionary* bzw. Alltags-Ontologien wie *ConceptNet*). Dass Cluster-Verfahren und *topic models* oder alternative Klassifikationsverfahren (CRFs, SVMs usw.) in seinem Methodenrepertoire keinerlei Rolle spielen, überrascht nicht so sehr, da er ein Apologet einfacher, unmittelbar interpretierbarer Verfahren ist (sofern sie die von ihm hypothetisch angenommen Trends auch abbilden können).

Aber eine Betonung solcher Mängel an methodischer Diversifikation wird dem eigentlichen Anliegen des Buches nicht gerecht. Und dies ist das Plädoyer für modellgestütztes, quantitatives Arbeiten in einem wissenschaftlichen Umfeld, das weder die empirische Methodik noch deren computationale Verfahren und Techniken bislang nennenswert zur Erkenntnisgewinnung heranzieht. Dass dies auch in einem literaturwissenschaftlichen Umfeld bedenkenswert sein könnte, zeigt der Autor auf verdienstvolle Weise. Er fängt zudem den geisteswissenschaftlich geprägten Leser mit genuin literaturwissenschaftlichen Fragestellungen ein und entwickelt mit seinem dort unüblichen Instrumentarium solide empirische Evidenzen, die eine Reihe von Standardannahmen zu Genres, literarischem Prestige und der Rolle der Geschlechter erschüttern. Dass angesichts des zu erwartenden Lesepublikums und seiner methodischen Hintergründe zentrale (logistische Regression) und am Rande erwähnte technische Begriffe (Rangfolgenkoeffizient, Signifikanzwert, Konfidenzintervall, Perzentil, Stichproben-Bias usw.) generell nicht im Fließtext tutoriell erklärt werden (alternativ wäre hier auch ein Appendix angeraten gewesen), macht das Buch nicht unverständlich, schmälert aber das genauere Verständnis für den dann auch technisch nachfragenden Leser.