

Rezension

Melanie Andresen (2024): *Computerlinguistische Methoden für die Digital Humanities. Eine Einführung für Geisteswissenschaftler:innen*. Tübingen: Narr Francke Attempto (= narr Studienbücher), 240 Seiten, € 26,99 ISBN 978-3-8233-8579-0

Martin Siefkes

Computerlinguistische Verfahren haben die Geisteswissenschaften schon lange erreicht und als ein spannender, sich rasch weiterentwickelnder Teil des Methodenarsenals etabliert. In den Medienwissenschaften werden sie dort relevant, wo Presse-, Plattform-, Werbe-, Serien-, Film- und Social-Media-Kommunikate und entsprechende Korpora nicht allein qualitativ-interpretierend, sondern auch mit Ansätzen des Text Mining, der automatischen Annotation und statistischen Modellierung erschlossen werden.

Melanie Andresens Studienbuch, das sich diesen Methoden widmet, erscheint in einem günstigen Moment: Chatbots und große Sprachmodelle haben in den letzten Jahren die computerlinguistische Auswertung sprunghaft verbessert und zugänglicher gemacht. Waren bis vor kurzem Programmierkenntnisse, ein (möglichst Linux-)Rechner mit allerlei Spezialtools und jahrelange Erfahrung *de rigueur* für ernsthafte Computerlinguist:innen, so ist heute für die Analyse multimodaler Kommunikate, sofern sie in digital(isiert)er Form vorliegen, Spezialsoftware nicht mehr zwingend erforderlich. Jede:r kann ein Bild oder eine gescannte Printmedienseite in einen Chatbot hochladen und dazu Fragen stellen oder ein paar Fachbegriffe eintippen: Schon wird ein leistungsfähiges KI-Modell, von denen viele längst multimodale Verarbeitungsfähigkeiten besitzen, auf das Analysat losgelassen. Gerade die Niedrigschwelligkeit dieses Computereinsatzes, im Verbund mit den äußerst voraussetzungsreichen dahintersteckenden Technologien, ist problematisch. Sie lässt es umso wichtiger erscheinen, zu verstehen und zu hinterfragen, wie sich computationale Methoden entwickelt haben, was sie mit den

Kontaktperson:

Dr. phil. habil. Martin Siefkes
Technische Universität Chemnitz
Thüringer Weg 11
09126 Chemnitz
martin.siefkes@phil.tu-chemnitz.de

analysierten Kommunikaten machen und wie ihre Anwendung den Blick formt.

Gleichzeitig besteht in vielen geistes- und auch medienwissenschaftlichen Studiengängen immer noch Nachholbedarf und Handlungsdruck. Lange Zeit bestand eine methodische Distanz zu den durch computationale Methoden ermöglichten Verfahren, Methoden und neuen Interpretationsansätzen, die gelegentlich auch in Abgrenzung zu den naturwissenschaftlich-technischen Fächern gezielt gepflegt wurde. Andresens Buch will diese – heute nicht mehr zeitgemäße – Distanz nicht durch ein weiteres Programmierlehrbuch überbrücken. Vielmehr will sie eine Einführung in computerlinguistische Grundbegriffe, Verfahren und Denkweisen leisten, die auf die spezifischen Fragestellungen, Methoden und Datentypen der Geisteswissenschaften zugeschnitten ist.

Der Band ist aus einer Stuttgarter Master-Lehrveranstaltung zu „Computerlinguistische[n] Methoden für die Digital Humanities“ hervorgegangen, die Andresen zwischen den Wintersemestern 2020/21 und 2023/24 unterrichtete (S. 9). Diese Genese merkt man dem Buch an, und zwar im guten Sinne: Die Kapitel sind didaktisch klar aufgebaut, beginnen meist mit einer knappen Orientierung, führen Grundbegriffe sorgfältig ein, zeigen Anwendungsfälle und schließen mit Übungen. Der Band richtet sich an Leser:innen ohne linguistisches, mathematisches oder programmiertechnisches Vorwissen (S. 11). Dass dennoch auch maschinelles Lernen, Deep Learning und Transformer-Modelle aufgenommen werden, macht den Anspruch aus: Es geht nicht nur um Korpusabfragen oder Verwendung von Tools, sondern um ein methodisches Verständnis der Prozesse von simplen Frequenzanalysen bis hin zu komplexen NLP-Verfahren.

Kontext und Abgrenzung

Im deutschsprachigen Feld der Digital Humanities schließt Andresens Einführung eine erkennbare Lücke. Der von Fotis Jannidis, Hubertus Kohle und Malte Rehbein herausgegebene Band *Digital Humanities. Eine Einführung* (Jannidis et al. 2017) bietet einen breit angelegten Überblick über Begriffe, Datenmodelle, Digitalisierung, digitale Objekte, Methoden sowie Fragen von Recht und Ethik. Er ist für die Digital Humanities einschlägig, aber nicht als systematische Einführung in computerlinguistische Methoden konzipiert, zumal Digital Humanities sich auch aus anderen Quellen speist. Andresens gemeinsam mit Heike Zinsmeister verfasster Band *Korpuslinguistik* (2019) ist näher am Gegenstand, bleibt als knapper narr-STARTER

jedoch stärker auf Korpora, Suchverfahren, Korpusaufbau und Annotation bezogen. Martin Weißers *Python-Programmierung für Germanist:innen* (2022), ebenfalls bei Narr erschienen, ergänzt den Werkzeugkasten in anderer Richtung: Dort steht die im DH-Kontext dominante Programmiersprache Python im Mittelpunkt, während Andresen gerade nicht primär Programmierkompetenz, sondern methodisches Verständnis und Urteilskompetenz vermittelt. Englischsprachige NLP-Einführungen und computerlinguistische Standardwerke wiederum sind häufig technischer, stärker mathematisch und zu wenig auf geisteswissenschaftliche Forschungslogiken zugeschnitten. Andresens Beitrag besteht also weniger in der Präsentation neuer Methoden als in der Übertragungsleistung zwischen Computer- und Korpuslinguistik sowie Statistik und KI-Terminologie und deren sich rasch erweiternden Anwendungsfeldern. Gerade in dieser Übertragung liegt der Mehrwert dieser Einführung – nicht nur, aber gerade auch für Medienlinguist:innen.

Gang durch die Kapitel

Das einleitende **Kapitel 1** führt in Gegenstand, Aufbau und Zielsetzung ein. Andresen beginnt bei alltagsnahen Beispielen computerlinguistischer Anwendungen – Suchmaschinen, automatische Übersetzung, Rechtschreibkorrektur, Wortvorschläge, Sprachassistentz, Chatbots – und verschiebt den Blick dann auf geisteswissenschaftliche Textanalyse. Distant Reading wird als Umgangsweise mit großen Textmengen eingeführt, das traditionelle Interpretation nicht ersetzt, sondern dieser vorgelagert wird. Nützlich ist die Abgrenzung zwischen Korpus- und Computerlinguistik.

Kapitel 2 behandelt die Lexik. Andresen unterscheidet Lexem, syntaktisches Wort, graphematisches Wort sowie Type und Token und führt daran die Notwendigkeit computerlinguistischer Modellierung vor. Es folgen Tokenisierung und Lemmatisierung als basale Vorverarbeitungsschritte, bevor das Kapitel zu allgemeinen Frequenzeigenschaften von Korpora, Kollokationen und Keywords weitergeht. Die Beispielstudien reichen von Fußballsprache über Coca-Cola-Werbung und Lake-District-Reiseberichten bis zu Wahlprogrammen.

Kapitel 3 widmet sich den Wortarten. Andresen rekonstruiert zunächst linguistische Kriterien der Wortartenbestimmung und erläutert dann, wie Wortarten annotiert und automatisch getaggt werden können. Dabei kommen nicht nur Tagsets und POS-Tagging zur Sprache, sondern auch Fehler, Grenzen und die Abhängigkeit automatischer Verfahren von Trainingsdaten und Domänen.

Kapitel 4 erweitert die Perspektive durch Einbeziehung der Syntax. Es führt Konstituenten- und Dependenzgrammatik ein und fragt, wie syntaktische Strukturen computerbasiert analysiert werden können. Der Gewinn gegenüber rein oberflächenorientierten Verfahren liegt darin, dass die Erfassung syntaktischer Relationen innerhalb von Sätzen Zugang zu dargestellten Handlungsmustern, Verhältnissen zwischen Personen oder Eigenschaftszuschreibungen bieten kann.

Kapitel 5 führt über das Konzept der Wortfelder in die Semantik ein. Nach knappen linguistischen Grundlagen zu Bedeutung, Denotation, Konnotation und semantischen Relationen geht es um Ressourcen und Verfahren, mit denen thematische Felder erstellt und auf Korpora angewendet werden können. Andresen zeigt das Potenzial solcher Ansätze anhand von korpusbasierten Studien zu deutschsprachigen Dramen, britischen Romanen des 19. Jahrhunderts, medialer Berichterstattung über Verteidigungsausgaben sowie deutschsprachigen Verschwörungsdiskursen.

Kapitel 6 behandelt Sentiment- und Emotionsanalyse. Andresen unterscheidet lexikonbasierte Sentimentanalyse, maschinelles Lernen und weitergefasste Formen der Emotionsanalyse. Sie erläutert, wie Sentimentwörterbücher entstehen, warum Kontext, Negation, Ironie, Metaphern und Domänenabhängigkeit große Probleme machen und weshalb automatisierte Bewertungsklassifikation nicht einfach mit menschlicher Interpretation gleichzusetzen ist. Die Beispielstudien führen von Happy-End-Erkennung in Romanen über Emotionsverläufe in Dramen und emotionale Kartierungen Londons bis zu Kriegstagebüchern und Pressemitteilungen politischer Parteien. Für eine medienlinguistische und politikwissenschaftliche Rezeption ist dieses Kapitel besonders anschlussfähig, weil Bewertung, Affekt und Emotionalisierung zentrale Kategorien öffentlicher Kommunikation sind. Zugleich macht Andresen deutlich, dass Sentimentanalysen anfällig für Kurzschlüsse werden, wenn Skalenwerte ohne differenzierte Texttyp- und Genremodelle abgelesen werden.

Kapitel 7 ist der distributionellen Semantik gewidmet. Ausgehend vom Grundsatz, dass Wörter mit ähnlicher Bedeutung in ähnlichen Kontexten auftreten, stellt Andresen Ähnlichkeitsmaße, Vektorräume und Word Embeddings vor. Sie erläutert spärliche und dichte Repräsentationen, statische und dynamische Embeddings und diskutiert Nutzungsmöglichkeiten sowie Evaluation. Die Stärke des Kapitels liegt darin, mathematisch anspruchsvolle Konzepte ohne Formalisierung zugänglich zu machen. Für Mediengeschichte und Diskursanalyse ergeben sich spannende Möglichkeiten, Bedeutungswandel oder semantische Nachbarschaften auf großen Textmengen zu untersuchen und damit Hypothesen über veränderte Konzeptualisierungen zu prüfen.

Kapitel 8 verlagert den Fokus auf Pragmatik und Referenz. Named Entity Recognition, Koreferenz und Named Entity Linking werden als Verfahren eingeführt, mit denen Texte über Personen, Orte, Organisationen, Werke und andere Entitäten erschließbar werden. Andresen betont, dass Entitäten referenzielle Anker von Texten bilden und deshalb für viele geisteswissenschaftliche Fragestellungen produktiv sind. Die Beispiele reichen von Figuren- und Netzwerkstudien zu Artusromanen und Wertheriaden über das Briefkorpus *Dehmel digital* bis zu historischen Zeitungsdaten und Karrierewegen von Migrant:innen. Für die Medienlinguistik liegt der Nutzen auf der Hand: Wer die Akteurskonstellationen öffentlicher Diskurse, die Sichtbarkeit von Institutionen oder die räumliche Semantik von Berichterstattung anhand von Textkorpora untersuchen will, findet hier das methodische Fundament.

Mit **Kapitel 9** beginnt der methodenorientierte zweite Teil. Korpusuche und Statistik werden über reguläre Ausdrücke, absolute und relative Frequenzen, deskriptive Statistik, Visualisierung und Inferenzstatistik eingeführt, deren zentrale Wichtigkeit für die Zuverlässigkeit von Ergebnissen hervorgehoben wird. Gerade für geisteswissenschaftliche Kontexte ist wichtig, dass Statistik hier nicht als Autoritätsgestus erscheint, sondern als Verfahren der kontrollierten Beschreibung und vorsichtigen Generalisierung.

Kapitel 10 behandelt manuelle Annotation und deren Voraussetzungen. Es gehört zu den zentralen Kapiteln des Bandes, weil es die Brücke zwischen hermeneutischer Kategorienbildung und automatisierbarer Datenarbeit explizit macht. Andresen erklärt, wann manuell annotiert werden sollte, wie Annotationsrichtlinien entstehen, wie Qualität und Reliabilität über Inter-Annotator-Agreement und Cohens Kappa geprüft werden können und welche Tools sich für unterschiedliche Aufgaben eignen.

Kapitel 11 führt in maschinelles Lernen ein, was zunächst nach einem großen Schritt im Anspruchsniveau klingt. Doch auch hier bleibt die Einführung gut lesbar. Andresen grenzt regelbasierte Systeme von lernenden Verfahren ab, unterscheidet überwachtes und unüberwachtes Lernen und beschreibt einen Musterablauf der Klassifikation: Trainingsdaten, Merkmale, Lernverfahren, Evaluation. Der klassische Naive-Bayes-Algorithmus wird exemplarisch durchgerechnet; Evaluation wird über Konfusionsmatrix, Accuracy, Precision, Recall und F1-Score erläutert. Der didaktische Gewinn liegt darin, dass maschinelles Lernen nicht als magische Black Box erscheint, sondern als Abfolge von Entscheidungen, die von Datenqualität, Kategorienbildung, Merkmalswahl und Evaluationsdesign abhängen.

Kapitel 12 setzt diese Linie mit Deep Learning konsequent fort. Künstliche neuronale Netze, Schichten, Gewichte, Verlustfunktionen, Training, Word Embeddings, Recurrent Neural Networks und Transformer werden in Grundzügen erläutert. Andresen betont sowohl die Leistungsgewinne als auch die Nachteile: große Datenmengen, Rechenaufwand, geringe Interpretierbarkeit und die Schwierigkeit, Modelle selbst zu trainieren. Praktische Hinweise auf Keras, PyTorch, Hugging Face und Google Colab ordnen ein, wie heutige Modelle nachgenutzt oder feinjustiert werden können. Das Kapitel ist notwendigerweise komprimiert. Wer tatsächlich Deep-Learning-Modelle bauen will, braucht weitere Literatur und entweder Programmiererfahrung oder den Mut, sich von aktuellen KI-Modellen (die bereits gut programmieren können) Schritt für Schritt durch die Aufgabe helfen zu lassen. Für den Zweck des Buches ist aber entscheidend, dass Leser:innen die Grundprinzipien, Voraussetzungen und Grenzen transformerbasierter Sprachverarbeitung einordnen können.

Kapitel 13 schließt mit Computerlinguistik und Ethik. Ausgehend von GPT-2 und der Debatte um die Veröffentlichung großer Sprachmodelle diskutiert Andresen Dual Use, Bias und Diskriminierung, Ursachen von Bias, Ressourcenverbrauch und Repräsentation. Das Kapitel ist kurz, aber wichtig, weil es ethische Fragen als Bestandteil methodischer Kompetenz behandelt. Für die Medienlinguistik sind diese Abschnitte direkt anschlussfähig: Automatische Sprachverarbeitung ist nicht nur ein Werkzeug zur Analyse von Medien, sondern selbst Teil medialer Infrastrukturen, Plattformpolitiken, Sichtbarkeitsordnungen und Diskriminierungsdynamiken. Dass Andresen auch ökologische Kosten und Fragen der Repräsentation aufnimmt, erweitert den Blick über reine Modellierungsfragen hinaus.

Leistungen und Stärken

Die größte Leistung des Buches ist seine didaktische Balance. Andresens Zielgruppe sind Geisteswissenschaftler:innen mit methodologischem Grundwissen, aber ohne Programmierkenntnisse oder besondere Technikaaffinität – all diejenigen, die nicht schon längst auf den Digital-Humanities-Zug aufgesprungen sind, die Relevanz computerlinguistischer Verfahren jedoch erkennen und ihr Verständnis und Methodenarsenal erweitern wollen. Andresen trifft die Bedürfnisse dieser Zielgruppe. Sie bietet eine zugängliche, aber wissenschaftlichen Ansprüchen genügende Darstellung und verbindet dies mit einem klaren didaktischen Aufbau (Erläuterungen an Praxisbeispielen und Korpora, gut machbare Übungen). Sie verzichtet

auf technische Virtuosität als Selbstzweck und achtet stattdessen auf Zugänglichkeit und Praxis: Was macht ein Verfahren? Welche Daten braucht es? Welche Annahmen liegen ihm zugrunde? Welche Fehler sind erwartbar? Welche Ergebnisform ist interpretierbar? Diese Fragen sind für geistes- und medienwissenschaftliche Forschung oft wichtiger als die Fähigkeit, jedes Verfahren selbst zu implementieren oder die Formeln dahinter zu verstehen.

Eine zweite Stärke ist die konsequente Verbindung von linguistischen und allgemein geisteswissenschaftlichen Grundlagen mit konkreten Verfahren und Anwendungsbeispielen. Dabei wird die prinzipielle Problematik, Kategorien für computationale Methoden zu „operationalisieren“, nicht ausgeblendet. Die Kapitel zu Lexik, Wortarten, Syntax, Semantik und Referenz zeigen jeweils, dass computerlinguistische Kategorien nicht bloß technische Vorstufen sind, sondern epistemologische Entscheidungen: Wer tokenisiert, lemmatisiert, taggt, parst, Entitäten erkennt oder Bezüge modelliert, legt fest, welche Einheiten und Relationen im Material überhaupt erst sichtbar werden. Das ist nicht zuletzt für die Medienlinguistik relevant, weil ihre Gegenstände überwiegend keine in Experimenten erhobenen „Daten“ sind, sondern in mediale, institutionelle und historische Kontexte eingebettete Kommunikate, die in komplexen Abhängigkeitsrelationen stehen. Durch die Digitalisierung und „Datifizierung“ wird diesen Kommunikaten und ihrer lebensweltlichen Bedeutung von Anfang an Gewalt angetan; schon die Idee einer repräsentativen Stichprobe, die jeder Inferenzstatistik zugrunde liegt, ist dabei eine Fiktion (oder bestenfalls eine Annäherung). Anzumerken ist jedoch, dass diese epistemische Gewalt, bei aller Vorsicht, nicht gänzlich vermeidbar ist und die Alternative, durch Verzicht auf die computationalen Methoden ein ganzes Spektrum an wissenschaftlichen Fragestellungen wieder der Beantwortung zu entziehen, wissenschaftstheoretisch betrachtet kaum akzeptabel erscheint. Auch wenn Andresen diese Abwägungen und eine durchaus mögliche kritisch-einordnende Perspektive auf quantifizierende Methoden nicht sehr tiefgehend diskutiert, ist eine solche Perspektive mit der Lektüre des Buchs und der Berücksichtigung seiner methodischen Anleitungen vereinbar.

Drittens geht der Band erfreulich differenziert mit Automatisierung um. Andresen argumentiert weder technikbegeistert noch technikskeptisch. Computerlinguistische Verfahren werden als nützlich, aber fehleranfällig; gut skalierend, aber reduktiv; leistungsfähig, aber interpretationsbedürftig dargestellt. Diese Haltung unterscheidet das Buch von vielen Einführungen, die entweder vor allem Tools präsentieren oder aktuelle KI-Verfahren als unausweichli-

chen Fortschritt rahmen. Überzeugend ist die durchgehende Aufmerksamkeit auf Evaluation, Datenqualität, Domänenanpassung und Interpretierbarkeit.

Viertens erhöht die praktische Ausstattung den Nutzwert. Übungen, digitale Musterlösungen, Python-Skripte (im digitalen Anhang), ein Ressourcenverzeichnis und ein Sachregister erhöhen sicherlich die praktische Einsatzbarkeit des Bands in der Lehre, sind aber auch fürs Selbststudium nützlich. Das Foodblogkorpus als wiederkehrendes Beispiel schafft ein gewisses Maß an Kontinuität. Zugleich werden aber auch viele externe Tools, Ressourcen und Korpora benannt, sodass Leser:innen nicht bei abstrakten Konzepten stehen bleiben, wobei diese Angaben bereits jetzt, zwei Jahre nach Erscheinen, eher als Ausgangspunkt für eine eigene Websuche betrachtet werden sollten. Zu empfehlen ist das Buch uneingeschränkt – jedoch nicht unergänzt – insbesondere für Einführungs- und Methodenseminare in allen geisteswissenschaftlichen Fächern.

Schwächen und Grenzen

Die Grenzen des Buches ergeben sich aus seinem Zuschnitt. Der Untertitel „für Geisteswissenschaftler:innen“ verspricht eine weite Ausrichtung, tatsächlich ist der Band jedoch vor allem eine Einführung in sprachbezogene Verfahren für textbasierte Geisteswissenschaften. Diese Schwerpunktsetzung auf schriftsprachliche Korpora und Analysemethoden mag der klassischen Computerlinguistik gerecht werden, ist jedoch heute nicht mehr ganz zeitgemäß. Für medienwissenschaftliche Fragestellungen ist sie besonders limitierend, da diese naturgemäß multimedial bzw. multimodal zu fassen sind. Audiovisuelle, visuelle, auditive, webbasierte oder interaktive Gegenstände – Filme, Videos, Bilder, Webseiten, Plattformmaterialien, Computerspiele – erfordern andere Datenmodelle und Annotationspraktiken, teils auch andere Auswertungsverfahren und Reflexionsrahmen. Für diese Erweiterung liegen inzwischen einschlägige Anschlussarbeiten vor, etwa das medienwissenschaftlich ausgerichtete *Handbuch Digitale Medien und Methoden* (Niebling et al. 2025), das von Arnold/Tilton (2023) ausgearbeitete Konzept des *Distant Viewing* für computergestützte Bild- und Videoanalyse, sowie zahlreiche Arbeiten der Digital Film Studies sowie methodische Grundlagen aus Web History, Digital Sound Studies und Game Research Methods. Gerade im Vergleich zu diesen Ansätzen wird deutlich, dass Andresens Einführung kein allgemeines Methodenbuch zur digitalen Medien- und Kommunikationsforschung sein will, sondern eine sprachdatenzentrierte Einführung in computerlinguistische Verfahren für geisteswissenschaftliche Anwendungen.

Auch der medienwissenschaftliche Theoriezusammenhang bleibt naturgemäß randständig. Zentrale Begriffe wie Medium, Öffentlichkeit, Plattform, Interface, Dispositiv, Zirkulation, Affordanz und viele weitere spielen keine tragende Rolle. Das ist kein Versäumnis im engeren Sinne, weil Andresen kein medienwissenschaftliches Methodenbuch schreibt. Für das Publikum des *Journals für Medienlinguistik* sei jedoch gesagt, dass hier die Anschlussleistung auf Seiten der Leser:innen liegt: Sie müssen die computerlinguistischen Verfahren in eigene Theorie- und Gegenstandslogiken übersetzen und die Anwendbarkeit auf fachspezifische Konzepte, und die sich aus diesen ergebenden Fragestellungen, im Einzelfall prüfen. Zugleich sei jedoch herausgestellt, dass die hier vorgestellten Methoden auch und gerade für die Medienwissenschaften, wo sie nicht durchgehend angekommen sind, eine große Bereicherung sind, wenn diese Adaption geleistet wird.

Eine weitere Grenze ist der Geschwindigkeit des Feldes geschuldet. Der Band erschien 2024, beruht auf Lehrveranstaltungen bis 2023/24 und nimmt Transformer sowie große Sprachmodelle auf. Dennoch kann er die seit ChatGPT nochmals beschleunigten Praktiken prompt-basierter Analyse, generativer Annotation, Retrieval-Augmented Generation, synthetischer Trainingsdaten, LLM-basierter Kodierung und deren methodische Validierung nur begrenzt behandeln. Zum jetzigen Zeitpunkt wäre daher – nach den Maßstäben anderer Felder – schon wieder eine Neuauflage denkbar. Das ist natürlich kein Fehler des Buches (zumal der Rezensent, wie hier offengelegt sei, mit der Rezension zeitnah nach Erscheinen des Bandes beauftragt wurde). Vielmehr handelt es sich um ein strukturelles Problem gedruckter Einführungen in ein dynamisches Feld. Positiv ist, dass Andresen über weite Strecken Grundlagen vermittelt, die auch nach der nächsten Tool-Generation oder dem nächsten Modellwechsel relevant bleiben. Wer konkrete Workflows für gegenwärtige LLM-gestützte Medienanalyse entwickeln will, informiert sich zusätzlich in aktuellen Tutorials oder Methodenblogs.

Schließlich hätte ein durchgehendes Projektbeispiel den Transfer noch erleichtern können. Das Foodblogkorpus sorgt zwar für Wiedererkennbarkeit, doch die Vielzahl der Methoden bleibt weitgehend kapitelweise organisiert. Ein abschließendes Kapitel, das ausgehend von einer Forschungsfrage den gesamten Prozess von Korpusaufbau, Metadaten, Annotation, Vorverarbeitung, Analyse, Evaluation, Visualisierung und Interpretation exemplarisch zusammenführt, hätte den Band noch gestärkt. Für Lehrende lässt sich ein solches Projektbeispiel jedoch mit etwas zusätzlichem Aufwand auf Basis der vorhandenen Übungen entwickeln. Voraussetzung dafür

ist jedoch ein Grundverständnis der Methoden, ihres Anwendungsspektrums und ihrer Fallstricke. Dafür ist das vorhandene Buch besonders geeignet.

Fazit

Andresens *Computerlinguistische Methoden für die Digital Humanities* ist ein sehr empfehlenswertes Studienbuch für alle, die sprachbezogene digitale Methoden verstehen und reflektiert in geisteswissenschaftliche Forschung integrieren wollen. Seine besondere Stärke liegt nicht in technischer Tiefe, sondern in der verständlichen und für Geisteswissenschaftler:innen relevant gehaltenen Einführung in Voraussetzungen, Begriffe, Verfahren und Grenzen computerlinguistischer Textanalyse. Andresen gelingt es, zwischen linguistischer Präzision, computerlinguistischer Methodik und geisteswissenschaftlicher Interpretationspraxis die Mitte zu finden.

Für medienwissenschaftliche Leser:innen ist der Band vor allem dann gewinnbringend, wenn sie Fragestellungen an ihre Gegenstände richten, bei denen Sprache eine mehr oder minder große Rolle spielt. Dies wird etwa bei der Untersuchung von Presse- und Plattformdiskursen, Werbe- und PR-Kommunikation, audiovisuellen Paratexten, Transkripten, Social-Media-Kommunikation, Kritiken, Serien- und Filmreviews, politischer Kommunikation, historischen Medienarchiven, Videountertitelungen und so manchen weiteren Formaten der heutigen Medienlandschaft häufig der Fall sein. Dabei wird kein fertiges Methodendesign geliefert, der Band will auch nicht zur Theorie digitaler Medien beitragen oder die Methodologie ihrer Untersuchung grundlegend updaten. Er vermittelt vielmehr ein solides methodisches Fundament, um computerlinguistische Verfahren nicht nur anzuwenden, sondern ihre Ergebnisse, Fehler, impliziten Annahmen und Erkenntnischancen beurteilen zu können. Gerade in einer Zeit, in der KI-Systeme als Werkzeuge und Gegenstände der Medienwissenschaften gleichermaßen präsent sind, ist dieser Durchblick, der in der Vielfalt der Tools schnell verloren geht, eine zentrale Kompetenz.

Literatur

- Andresen, Melanie/Zinsmeister, Heike (2019): *Korpuslinguistik*. Tübingen: Narr Francke Attempto (= narr Starter).
- Arnold, Taylor/Tilton, Lauren (2023): *Distant Viewing: Computational Exploration of Digital Images*. Cambridge, MA: MIT Press.

- Jannidis, Fotis/Kohle, Hubertus/Rehbein, Malte (Hg.) (2017): *Digital Humanities. Eine Einführung*. Stuttgart: J. B. Metzler.
- Niebling, Laura/Raczkowski, Felix/Stollfuß, Sven (2025): *Handbuch Digitale Medien und Methoden*. Wiesbaden: Springer VS.
- Weißer, Martin (2022): *Python-Programmierung für Germanist:innen. Ein Lehr- und Arbeitsbuch*. Tübingen: Narr Francke Attempto (= narr Studienbücher).