

Henning Lobin, Roman Schneider & Andreas Witt (Hg.). 2018. *Digitale Infrastrukturen für die germanistische Forschung* (Germanistische Sprachwissenschaft um 2020, Band 6). Berlin, Boston: De Gruyter. 373 S.

Besprochen von **Karina Frick**: Universität Zürich, Schönberggasse 9, CH-8001 Zürich, E-Mail: karina.frick@uzh.ch

<https://doi.org/10.1515/zrs-2020-2047>

Der zu rezensierende Sammelband beschließt die sechsbändige Reihe *Germanistische Sprachwissenschaft um 2020*, die anlässlich Ludwig M. Eichingers Eintritt in den Ruhestand erschienen ist. Die von Albrecht Plewnia und Andreas Witt herausgegebene Reihe spannt in sechs thematisch abgeschlossenen Bänden ein Panorama über aktuelle Themengebiete der germanistischen Sprachwissenschaft auf; der letzte Band widmet sich nun Ludwig M. Eichingers beträchtlichen – und in vielerlei Hinsicht visionären – Verdiensten um digitale Forschungsinfrastrukturen in der Sprachwissenschaft im Rahmen seiner langjährigen Tätigkeit als Direktor des Leibniz-Instituts für Deutsche Sprache.

Bevor auf den Inhalt der einzelnen Beiträge eingegangen wird, kann zunächst festgehalten werden, dass der Band insgesamt geprägt ist von der „Vision einer vollständigen Interoperabilität sämtlicher Forschungsdaten mit all ihren Metadaten“ (S. 6) sowie der damit einhergehenden Forderung nach Kooperativität; das schlägt sich unter anderem darin nieder, dass trotz oder vielleicht gerade wegen der sehr großen Bandbreite an Beiträgen relativ viele inhaltliche Redundanzen vorliegen, was beispielsweise die Definition des Begriffs *Infrastruktur* (vgl. z.B. S. 54 und S. 252)¹ betrifft, aber auch die wiederholten Plädoyers für Standardisierung (auf unterschiedlichen Ebenen). Insbesondere Letzteres kann allerdings durchaus auch positiv gewertet werden: Trotz der zum Teil sehr unterschiedlichen wissenschaftlichen Provenienzen der Autor*innen scheint – zumindest ansatzweise – Konsens darüber zu bestehen, in welche Richtung sich die germanistische Forschung im Zuge der Digitalisierung im Rahmen der titelgebenden *digitalen Infrastrukturen* bewegen sollte.

1 Auf der anderen Seite aber wird der Begriff in den einzelnen Beiträgen zum Teil sehr unterschiedlich verstanden; insbesondere die Abgrenzung zu Sprachressourcen erschließt sich dabei nicht immer. Hier hätte ein rahmengebendes Einführungskapitel womöglich Abhilfe schaffen können, wobei die Vielfalt der Beiträge wiederum darauf schließen lässt, dass das gerade nicht das Ziel der Herausgeber war.

Aufbau und Inhalt

Vier Großkapitel gliedern den Band: Kapitel I befasst sich mit „Kooperationen und Verbänden“ und beleuchtet „die gegenwärtige Situation im Bereich von Forschungsinfrastrukturen und -ressourcen“ (S. 2). In Kapitel II stehen „Sprachwissenschaft und Sprachtechnologie“ im Fokus; es soll also die Nutzung der im ersten Kapitel beschriebenen Infrastrukturen mit Bezug zu spezifischen sprachwissenschaftlichen Forschungsfragen in den Blick genommen werden. Kapitel III widmet sich „Korpora und Informationssystemen“, wobei Entstehung und Eigenschaften exemplarischer Korpora und Informationssysteme vorgestellt werden. Thema des letzten Kapitels IV schließlich sind „Annotation und Modellierung“. Darin geht es um die Voraussetzungen der kooperativen Nutzung von Daten durch standardisierte Verfahren und Beschreibungsregeln. Im Folgenden werden nun die einzelnen Beiträge der Kapitel jeweils kurz zusammengefasst und das betreffende Kapitel als Ganzes beurteilt.

Der erste Beitrag in Kapitel I stammt von THOMAS GLONING. Er befasst sich mit den weitreichenden Veränderungen durch die Digitalisierung in der germanistischen Forschung und Lehre. Diese betreffen nicht nur die Verfügbarkeit und den adäquaten Umgang mit digitalen Daten und Werkzeugen, sondern auch „neue Formen der Zusammenarbeit“ (S. 12) und der „Vergemeinschaftung“ (S. 14). Der Beitrag schließt mit der Formulierung von sechs zukünftigen Aufgabenbereichen, die im Hinblick auf den Ausbau von Infrastrukturangeboten aus Sicht der Nutzer*innen relevant sind. ERHARD HINRICHS Beitrag fokussiert auf „Digitale Forschungsinfrastrukturen für die Sprachwissenschaft“ und lenkt den Blick somit auf fachspezifische Erfordernisse. Wie solche Forschungsinfrastrukturen den Anforderungen von Sprachdaten „in einer zuvor nie gekannten Größenordnung“ (S. 37) gerecht werden können, zeigt Hinrichs anhand des europäischen Verbundprojekts CLARIN auf. Dem zweiten großen europäischen Forschungsverbund für Geistes- und Sozialwissenschaften, DARIAH, widmet sich der Beitrag von STEFAN SCHMUNK, FRANK FISCHER, MIRJAM BLÜMM und WOLFRAM HORSTMANN. Einer historischen Einordnung von Forschungsinfrastrukturen im Allgemeinen folgt die Vorstellung von DARIAH sowie die Ableitung von Prinzipien für den Aufbau von Forschungsinfrastrukturen. Im letzten Beitrag des ersten Kapitels beleuchten KARLHEINZ MÖRTH und TANJA WISSIK anhand zahlreicher Beispiele „Digitale Sprachressourcen in Österreich“, wobei es sich im Wesentlichen um eine teilweise etwas kontextfreie Auflistung verschiedener Sprachressourcen handelt (deren Beschreibung womöglich mit der Publikation des Bandes bereits veraltet sind). Eine vertiefende Einordnung bzw. die Darstellung von Verknüpfungen wäre hier wünschenswert gewesen, zumal das Kapitel ja „Kooperationen und Verbände“, nicht aber Einzelbeispiele erwarten lässt.

Insgesamt vermag das Kapitel aber zu halten, was der Titel verspricht: Es handelt sich um einen Überblick inklusive historischem Abriss über die germanistische (aber v.a. sprachwissenschaftliche) Landschaft digitaler Infrastrukturen. Eine solche Übersicht hat als Dokumentation des status quo sicherlich ihre Berechtigung – zumal die Reihe ein „Panorama der ‚Germanistischen Sprachwissenschaft um 2020‘“ (S. VII) bieten möchte –, läuft allerdings insbesondere aufgrund der raschen Veränderlichkeit des behandelten Themas immer auch Gefahr, bei der Publikation bereits überholt zu sein.

Das zweite Kapitel zu „Sprachwissenschaft und Sprachtechnologie“ leiten HANNAH KERMES und ELKE TEICH mit einem Beitrag zu generischen Infrastrukturen und spezifischer Forschung sowie deren Zusammenspiel ein. Darin beschreiben sie zunächst Möglichkeiten und Vorteile generischer Infrastrukturkomponenten (z.B. vorhandene Korpora, Abfrage- oder Analysewerkzeuge) und zeigen anschließend anhand eines Beispiels deren Anpassung für spezifische Settings auf – also z.B. konkrete linguistische Fragestellungen oder eine spezielle Datengrundlage. Im Artikel werden außerdem methodische Aspekte beim Aufbau generischer Infrastrukturkomponenten reflektiert. Ebenfalls methodisch perspektiviert ist der Beitrag von KERSTIN ECKART, MARKUS GÄRTNER, JONAS KUHN und KATRIN SCHWEITZER, die sich sprachtechnologischer Infrastruktur widmen. Sie betonen die Relevanz technologischer Komponenten (z.B. die Gestaltung der Werkzeuginfrastruktur) für den gesamten Forschungsprozess, gehen dabei unter anderem auf die Vor- und Nachteile automatischer und manueller Annotation ein und zeigen auf, wie verschiedene Annotationsebenen (z.B. Prosodie und Syntax) zusammengeführt werden können. Der Beitrag enthält auffällig viele Orthographie- und Grammatikfehler (z.B. S. 126, S. 131, S. 133, S. 140, S. 142) sowie fehlerhafte Links in den Fußnoten auf S. 143, was in dem ansonsten bis dahin sauber redigierten Band negativ heraussticht. ALEXANDER MEHLER, WAHED HEMATI, RÜDIGER GLEIM und DANIEL BAUMARTZ befassen sich mit Entwicklungstendenzen und dem damit einhergehenden Anpassungsdruck auf Infrastrukturen in den Digital Humanities. Wie diesen Anforderungen begegnet werden könnte, beschreiben die Autor*innen in ihrem voraussetzungsreichen, stark informatikgeprägten Beitrag exemplarisch anhand dreier Informationssysteme und deren Potentialen. „Am Beispiel der Rezeptionsforschung“ gehen HANS-JÜRGEN BUCHER und PHILIPP NIEMANN auf Infrastrukturen und deren bisher noch kaum genutzte Möglichkeiten in der Medienforschung ein. Sie plädieren dabei für eine stufenweise Standardisierung der Sicherung und Dokumentation von Forschungsdaten, da dadurch produktive Synergien ermöglicht und die Forschungsqualität auf längere Sicht erhöht werden können.

Während das erste Kapitel sowohl inhaltlich als auch sprachlich in einem leicht lesbaren Duktus gehalten war, hebt sich das zweite durch seinen höheren Anspruch und Komplexitätsgrad deutlich davon ab. Das dürfte wohl einerseits

intendiert sein, geht es hier doch zum einen spezifisch um Sprachwissenschaft – das Kernthema der Reihe – und lässt sich andererseits bei einem so breiten Spektrum an disziplinären Perspektiven ohnehin kaum vermeiden. Gleichwohl zeigt sich darin auch eine Schwierigkeit des Bandes: Die große thematische und disziplinäre Varianz erschwert eine durchgängige (bzw. eine durchweg nachvollziehbare) Rezeption. Zwar ist das insofern leicht zu umgehen, als dass Beiträge sehr gezielt rezipiert werden können, allerdings bleibt dann der Gesamtüberblick – bzw. das Panorama, das als eines der Hauptziele der Reihe genannt wird – verwehrt.

Das dritte Kapitel beginnt mit einem Beitrag von RUXANDRA COSMA und MARC KUPIETZ und handelt „Von Schienen, Zügen und linguistischen Fragestellungen“. Mit der Schienenmetapher wollen die Autor*innen verdeutlichen, wie sich durch die Digitalisierung die „Transportbahnen [...] immer mehr vernetzen“ und eine „in der Linguistik mittlerweile staunenerregende Fahrgeschwindigkeit“ (S. 199) erlauben. Forschung ohne institutionelle Infrastrukturen ist zwar noch immer möglich, die steigenden Ansprüche an Validität und Nachvollziehbarkeit in der empirischen Linguistik machen sie aber zunehmend unentbehrlich. Das verdeutlicht auch der folgende Beitrag von ALEXANDER GEYKEN, MATTHIAS BOENIG, SUSANNE HAAF, BRYAN JURISH, CHRISTIAN THOMAS und FRANK WIEGAND, die das *Deutsche Textarchiv* als mittlerweile in CLARIN integrierte Forschungsplattform vorstellen. Deren Funktionalitäten reichen von der Erstellung, Annotation und Kuratation von Korpora über die kollaborative Arbeit an den Dokumenten zur Qualitätssicherung bis hin zur Bereitstellung verschiedener Analysewerkzeuge. ANDREA RAPP widmet sich im Folgenden „Digitale[n] Forschungsinfrastrukturen in der Germanistischen Mediävistik“. Ausgehend von Überlegungen zum Verhältnis synchroner und diachroner sprachwissenschaftlicher Ansätze und einer definitorischen Einordnung des Begriffs *Forschungsinfrastrukturen* gibt Rapp anhand konkreter Beispiele (z. B. von Quellen, Korpora und Lexika) einen Überblick über in der Mediävistik vorhandenen Infrastrukturen und dem damit einhergehenden Paradigmenwechsel. Zwar werden hier konkrete Beispiele für mediävistische Korpora genannt, dennoch mag sich der Artikel nicht so recht in das Kapitel einfügen; er wäre meines Erachtens aufgrund seines Überblickscharakters besser im ersten Teil aufgehoben gewesen (bspw. anstelle des Beitrages von Mörth und Wissik, der seinerseits besser in dieses Kapitel gepasst hätte). Der Beitrag von MARTINE DALMAS und ROMAN SCHNEIDER schließlich nimmt „Die grammatischen Online-Angebote des IDS aus Sicht der Germanistik im Ausland“ in den Blick. Dabei zeigen die Autor*innen anhand von zwei digitalen grammatischen Informationssystemen des IDS auf, welche Einsatzmöglichkeiten diese für die Auslandsgermanistik bieten – aber auch, welche Erwartungen damit verbunden sind und wo Fallstricke oder Verbesserungsmöglichkeiten bestehen.

Insgesamt wirkt das dritte Kapitel als Ganzes inhaltlich eher lose zusammenhängend; die inhaltlichen Schwerpunkte der Beiträge in diesem Kapitel scheinen z.T. sehr weit auseinanderzuliegen, gemeinsame Grundlagen gibt es kaum (wobei das zumindest in Teilen auch auf die anderen Kapitel zutrifft).

Das letzte Kapitel eröffnet mit C. M. SPERBERG-McQUEEN, Mitherausgeber der *Extensible Markup Language* (XML) und der Richtlinien der *Text Encoding Initiative* (TEI), ein Spezialist und Pionier von Auszeichnungssprachen. Er widmet sich in seinem Beitrag den „Kernideen der deskriptiven Textauszeichnung“ anhand von XML und demonstriert exemplarisch deren Bedeutung für die interoperable wissenschaftliche Beschäftigung mit Texten. Der Beitrag von MICHAEL BEIßWENGER eröffnet eine korpuslinguistische Perspektive auf die internetbasierte Kommunikation. Er zeigt darin nach einem Überblick über gegenwärtige Korpora zunächst auf, dass sich trotz des fortwährenden technologischen Wandels stabile Interaktionsformate gebildet haben (z.B. *Posts*) und geht anschließend der Frage nach, auf welche Weise diese standardisiert repräsentiert werden können. Dazu greift er auf die Encoding-Standards der TEI zurück, die als de-facto-Standard auch in der Sprachwissenschaft bereits etabliert sind und deshalb ein hohes Potential für Interoperabilität aufweisen. Mit ihrem Beitrag zum Einsatz von Topic-Modellen in der philologischen Forschung schließen GERHARD HEYER, GREGOR WIEDEMANN und ANDREAS NIEKLER den Band ab. Darin gehen sie zunächst theoretisch auf das statistische Verfahren des *Topic Modelling* ein und schildern anschließend anhand von konkreten Beispielen dessen Einsatzmöglichkeiten in der philologischen Forschung. Der Beitrag verdeutlicht die Potentiale und Grenzen der beschriebenen Methode und plädiert überzeugend für den Einsatz solcher Verfahren in der germanistischen Forschung.

Während der erste Beitrag dieses Kapitels allgemein gehalten ist und eher Grundlagenwissen zu Auszeichnungssprachen vermittelt, sind die folgenden zwei umso spezifischer, entweder in Bezug auf die Datenbasis (Beißwenger) oder die verwendete Methode (Heyer/Wiedemann/Niekler). Auch dieses Kapitel ist somit geprägt von einer großen Varianz, die – wie weiter oben beschrieben – mit gewissen rezeptiven Herausforderungen verbunden ist; nichtsdestotrotz sind besonders die letzten beiden Artikel insofern wertvoll, als sie anhand spezifischer Kontexte die Potentiale methodischer Herangehensweise an Sprachressourcen aufzeigen und somit die teilweise theoretisch abstrakt bleibenden oder aber sehr (fall)beispielhaften Ausführungen durch eine anwendungsorientierte Perspektive ergänzen.

Fazit

Abschließend ist festzuhalten, dass der Band einige Schwächen aufweist, die allerdings auch auf das ambitionierte Ziel der Reihe zurückzuführen sein dürften. Störend sind etwa die zahlreichen Redundanzen (bspw. das wiederholte Porträtieren von CLARIN oder die wiederholten Plädoyers für Standardisierung). Zudem stellt sich bei der Lektüre des Bandes grundsätzlich die Frage nach der Sinnhaftigkeit der vorgenommenen Kapitelaufteilung; das mag indes auch daran liegen, dass die Überschriften insgesamt sehr allgemein gewählt und deshalb auch nur bedingt aussagekräftig sind. Die Unterteilung der Beiträge in vier Kapitel wirkt deshalb stellenweise etwas bemüht und konstruiert; sie vermag jedenfalls der großen thematischen Varianz der Beiträge nicht gerecht zu werden. Schließlich ist ebendiese Varianz mitunter auch herausfordernd, da die einzelnen Beiträge ganz unterschiedliches Vorwissen voraussetzen. Verdienstvoll hingegen ist der Band sicherlich insofern, als er durch die breitgefächerte und disziplinenübergreifende Zusammenstellung der Beiträge ganz unterschiedliche Perspektiven auf historische, gegenwärtige und mögliche zukünftige Wege der germanistischen Forschung ermöglicht. Insgesamt handelt es sich um eine inspirierende Lektüre, insbesondere mit Blick auf interdisziplinäre Synergiepotentiale, die – das macht der Band sehr deutlich – für zukünftige sprachwissenschaftliche Forschungsarbeit im Rahmen digitaler Infrastrukturen nicht nur wünschenswert, sondern geradezu unabdingbar sind.