

# Einsam oder gemeinsam?

## Verstehen und Erklären in der digitalen Linguistik

Marcus Müller 

Eingegangen: 31. Oktober 2023 / Angenommen: 30. Dezember 2023  
© The Author(s) 2024

**Zusammenfassung** Der Beitrag untersucht die Bedeutung der Unterscheidung zwischen Verstehen und Erklären für die digitale Linguistik. Es wird gezeigt, dass Verstehen und Erklären nicht als isolierte Praktiken betrachtet werden sollten, sondern eng miteinander verknüpft sind und sich gegenseitig ergänzen. Er setzt dazu bei der bekannten Begriffsdichotomie aus der romantischen Hermeneutik an und bezieht neuere Beiträge zu einem einheitswissenschaftlichen Begriff des Erklärens wie auch zu einer pragmatischen Verstehenstheorie ein. Das Verstehen wird in der digitalen Linguistik als eine unreflektierte Nebenbei-Praxis behandelt, während das Erklären in seinen verschiedenen Typen gut methodologisiert ist. Der in verschiedenen Bereichen der Linguistik eingeführte Kontextualisierungsbegriff scheint geeignet, um korpuslinguistisches Verstehen methodologisch zu fassen. Auch wenn die Bedeutung statistischer Erklärungen bei wachsenden Beständen an quantifizierbaren Daten wächst, darf das Verstehen im Kontext nicht unreflektiert bleiben und bedarf einer methodischen Basis. Auf der Ebene des praktischen Zugriffs hebt der Beitrag Annotation als Methode hervor, die Verstehen und Erklären integriert. Dabei spielt insbesondere das kollaborative Verstehen in Annotationsprozessen eine Schlüsselrolle. Der Beitrag endet mit zwei Anwendungsbeispielen.

**Schlüsselwörter** Verstehen · Erklären · Korpuslinguistik · Methodologie · Kontextualisierung · Annotation

---

✉ Marcus Müller  
Institut für Sprach- und Literaturwissenschaft, Technische Universität Darmstadt, Darmstadt,  
Deutschland  
E-Mail: [marcus.mueller@tu-darmstadt.de](mailto:marcus.mueller@tu-darmstadt.de)

## Lonesome or Together

Understanding and Explanation in Digital Linguistics

**Abstract** This paper examines the significance of the distinction between understanding and explanation for digital linguistics. It argues that understanding and explanation should not be regarded as isolated practices, but are closely linked and complement each other. To this end, it starts with the well-known conceptual dichotomy from Romantic hermeneutics and incorporates more recent contributions to a unified scientific concept of explanation as well as to a pragmatic theory of understanding. Understanding is treated in digital linguistics as an unreflective incidental practice, while explanations are well-methodologised in their various types. The notion of contextualisation, established in various linguistics fields, seems suitable to methodologically grasp corpus linguistic understanding. Even if the importance of statistical explanations grows with increasing stocks of quantifiable data, understanding in context must not remain unreflective and requires a methodological basis. On the level of practical access, the article highlights annotation as a method that integrates understanding and explanation. In particular, collaborative understanding plays a key role in annotation processes. The article ends with two examples of application.

**Keywords** Understanding · Explanation · Corpus Linguistics · Methodology · Contextualisation · Annotation

### 1 Einleitung

Die Natur erklären wir, das Seelenleben verstehen wir. (Dilthey [1894] 1990, S. 144)

Wilhelm Diltheys berühmtes Zitat bezieht sich bekanntlich auf den Methodenbruch zwischen erklärenden Naturwissenschaften und verstehenden Geisteswissenschaften in dem Sinne, dass Dilthey als fundamentales heuristisches Prinzip der Naturwissenschaften die Deduktion sieht. Damit ist die Ableitung von Erkenntnis aus einem absolut und allgemein gültigen Gesetz gemeint, nach dem am Einzelfall gesucht wird, während es in den Geisteswissenschaften um Sinnstiftung durch die empathische Ko-Konstruktion einer intentionalen Sinngestalt am einzelnen Sachverhalt geht. Die Unterscheidung zwischen erklärenden und verstehenden Wissenschaften hat Dilthey allerdings nicht erfunden, sie wird vielmehr auf den Historiker Johann Gustav Droysen (1868, S. 41–62) zurückgeführt. Dilthey hat sie aber in das Programm der Hermeneutik eingebettet. Verstehen in diesem Sinne heißt, eine geistige Hervorbringung in seiner Individualität zu erfassen und dazu den eigenen Erfahrungshorizont heranzuziehen.

In diesem Beitrag möchte ich die Dichotomie von Verstehen und Erklären auf die Methodendiskussion der digitalen Linguistik beziehen und daran einige methodische Punkte verdeutlichen, die mir beim Arbeiten mit digitalen Korpora wichtig

erscheinen.<sup>1</sup> Dazu greife ich zuerst Diltheys Begriff des Verstehens auf, allerdings in einer sehr punktuellen und zugespitzten Art und Weise (Kap. 2). Danach führe ich die Idee ein, das Erklären, also das Herausarbeiten allgemeiner Prinzipien, sei Aufgabe aller Wissenschaften und demnach auch der Geisteswissenschaften, so sie denn den Anspruch verfolgten, eine Wissenschaft zu sein. Dazu beziehe ich mich auf den Philosophen Gerhard Schurz, der sich wiederum vor allem mit dem Wissenschaftstheoretiker Carl Gustav Hempel auseinandersetzt (Kap. 3). Im Folgekapitel skizziere ich verschiedene Programme des Erklärens in der Linguistik und deren Rolle in der Korpuslinguistik (Kap. 4), um in Kap. 5 wieder auf das Verstehen zurückzukommen, diesmal aus der korpuslinguistischen Praxis heraus betrachtet, wo es mir als Nebenbeiprogramm im Windschatten der mit Regeln, Qualitätsmaßstäben und Lehrbüchern flankierten Methoden begegnet. Die Methode der Annotation scheint mir ein Ort zu sein, an dem das Verstehen und das Erklären in der Korpuslinguistik zusammengeführt und methodisch kontrolliert werden können. In Kap. 6 diskutiere ich das theoretisch Entfaltete an zwei Anwendungsbeispielen aus meiner eigenen Arbeit. Ich schließe mit einem Fazit (Kap. 7).

Die Vorstellungen zum Verstehen und Erklären entwickle ich in diesem Beitrag anhand meiner eigenen linguistischen Arbeit mit Korpora, die – mit Überschneidungen – in den Bereichen der lexikalischen Semantik, funktionalen Grammatik, Textpragmatik und Diskursanalyse angesiedelt ist. Natürlich gibt es Bereiche der Korpuslinguistik, auf die mein Konzept nicht, nicht ganz oder in anderer als der dargestellten Weise zutrifft. Ich bemühe mich im Folgenden, den intendierten Geltungsbereich meiner Aussagen deutlich zu machen, möchte diesen Absatz aber als salvatorische Klausel verstanden wissen, sollte sich doch noch Manches als unzulässige Verallgemeinerung lesen.

## 2 Verstehen als einsames Geschäft

Das Verstehen als geisteswissenschaftliche Methode hat selbstredend eine lange Begriffs- und Rezeptionsgeschichte, die schon oft erzählt worden ist und hier nicht nachvollzogen werden kann.<sup>2</sup> Hier sind nur die folgenden Punkte bedeutsam, die sich aus der Theoriewelt der romantischen Hermeneutik ergeben: Verstehen ist erstens eine Methode der Einfühlung und des geistigen Nachvollzugs (Dilthey [1894]

---

<sup>1</sup> Zur ›digitalen Linguistik‹ kann man Computerlinguistik, Korpuslinguistik, Natural Language Processing sowie Aspekte der Informationswissenschaft und der digitalen Editionswissenschaft zählen. Die Frage nach der Vollständigkeit dieser Aufzählung, die Beziehung der Begriffe zueinander und die Überlappungen ihrer Extensionen kann ich an dieser Stelle nicht diskutieren. Ich mache meine Ausführungen im Folgenden in einem hegemonialen Ausgriff an der Korpuslinguistik fest, behandle aber auch Praktiken, die in den anderen genannten Bereich relevant sind, z. B. Annotation und Segmentierung.

<sup>2</sup> Zum Begriff des Verstehens in der Philosophiegeschichte s. Apel (1955), in der Hermeneutik und pragmatischen Sprachphilosophie s. Scholz (2001), zum Verstehen aus Sicht der Sprachpsychologie s. Hörmann (1976), aus Sicht der linguistischen Hermeneutik s. Biere (1989) und die Beiträge in Hermanns/Holly (2007). Die Hinweise finden sich auch in Müller (2015, S. 10, FN 8). Hermanns (2009) erläutert das literarische Verstehen sehr nachdrücklich am eigenen Beispiel bei der Lektüre von Heinrich von Kleists *Anekdote aus dem letzten preußischen Kriege*. Liebert (2020) ergänzt Hermanns' Sichtweise um Einsichten zum Begriff der Empathie beim Verstehen.

1990, S. 318). Es geht also nicht um die Auswahl abstrakter Bedeutungspositionen im semasiologischen Feld sprachlicher Zeichen, sondern um den empathischen Nachvollzug dessen, was jemand mit der konkreten Disposition sprachlicher Zeichen im Text gemeint hat. Es ist der Nachvollzug einer geistigen Bewegung, der sozusagen durch die Zeichen hindurch geschieht, als wären sie ein Fenster in die Gedankenwelt einer schreibenden Person. Gadamer (1960, S. 296) nennt Verstehen das Einrücken in ein Überlieferungsgeschehen, in dem sich Vergangenheit und Gegenwart beständig vermitteln. Verstehen muss man sich zweitens als einen Sonderfall der zwischenmenschlichen Begegnung vorstellen. Da die Hermeneutik in ihren Hauptdomänen Theologie, Recht, Philosophie und Literatur und in jüngerer Zeit auch Linguistik immer eine Lehre und Theorie des Verstehens geschriebener Texte war, geht es um die Begegnung mit Abwesendem, und zwar in dem doppelten Sinne, dass sowohl die schreibende und meinende Person als auch das, worüber geschrieben wird, auf systematische Weise abwesend ist. Die Abwesenheit ist geradezu die Grundbedingung der Zeichenbildung. Verstehen in diesem Sinne ist eine doppelte Konstruktionsleistung, mit der das meinende Individuum und der gemeinte Sachverhalt unter den in einer Lese-Situation geltenden (historischen, weltanschaulichen, epistemischen) Bedingungen in eine semiotisch emulierte Ko-Präsenz gebracht werden. Das Verstehen als Ko-Konstruktion des Abwesenden kann und soll allerdings nur regelgeleitet vonstattengehen, das regelgeleitete Verstehen nennt Dilthey »Interpretation«:

Wir können auch solche Lebensäußerungen verstehen, die in Relikten, Zeichen, Symbolisierungen ihren materiellen Träger gefunden haben und somit über die Zeit hinweg erhalten bleiben. Die Fixierung der Lebensäußerung in einem materiellen, als Zeichen fungierenden Substrat stellt die Grundlage des geisteswissenschaftlichen Interpretationsbegriffs dar. Das durch Regeln disziplinierte ›Verstehen von dauernd fixierten Lebensäußerungen nennen wir Auslegung oder Interpretation.« (Teichert 2010, S. 22, der ein Zitat von Dilthey [1894] 1990, S. 319, aufnimmt)

Das Verstehen in diesem wissenschaftsprogrammatischen Sinne geht vom Individuum aus und richtet sich auf das Individuum.<sup>3</sup> Es ist ein einsames Geschäft. Teichert (2010, S. 21) drückt es folgendermaßen aus:

Das geisteswissenschaftliche Interesse gilt primär dem *individuellen* Erleben. Die kollektiven Bedingungen des Individuellen werden weder gezeugnet noch ignoriert. Aber sie stehen nicht im Zentrum des hermeneutischen Interesses. Dieses gilt der »wissenschaftlichen Erkenntnis der Einzelpersonen«.

Das hermeneutische Verstehen im Sinne der Rekonstruktion eines so-intendierten Sachverhaltes kann daher drittens nie vollständig gelingen und es kann auch nie vollständig von Anderen nachvollzogen werden – es bleibt immer individuell und

<sup>3</sup> Um das Individuelle des geisteswissenschaftlichen Verstehens zu betonen und gegen das Allgemeine der naturwissenschaftlichen Verfahren abzugrenzen, hat etwa zeitgleich mit Dilthey der Philosoph Wilhelm Windelband die »idiographische Methode der Geisteswissenschaft« von der »nomothetische[n] Methode der Naturwissenschaft« unterschieden (Windelband [1894] 1915, S. 136).

ist nicht replizierbar. Vor allem Letzteres wird man sich merken müssen, wenn es um die Bedingungen des Verstehens in der digitalen Linguistik geht (s. unten, Kap. 5).

### 3 Das Einheitsmodell der wissenschaftlichen Erklärung

Es hat gegen die Idee, jede akademische Kultur habe ihre je eigenen epistemischen Verfahren, die ihr eben gemäß sei, aber auch Einwände und Alternativprogramme gegeben, die im Gegenteil auf in allen Disziplinen geltende allgemeine Prinzipien des Erkennens abzielten. Schurz (2004) weist auf das Programm Carl Gustav Hempels hin, eines in den USA lehrenden Wissenschaftsphilosophen, der vom logischen Positivismus des Wiener Kreises beeinflusst war (vgl. Fetzer 2022).

Wie schon der Titel »The Function of General Laws in History« verrät, möchte Hempel in seiner ersten Arbeit zum Erklärungsbegriff von (1942)<sup>4</sup> entgegen dem Droysen-Dilthey-Windelbandschen Programm zeigen, daß die Suche nach allgemeinen Gesetzesmäßigkeiten ein unerläßlicher Schritt sei, auf den auch die Geisteswissenschaften und speziell die Geschichtswissenschaften *unvermeidlich* angewiesen sind, sobald sie etwas zu erklären bzw. zu verstehen suchen. Das sogenannte Modell der *deduktiv-nomologischen* Erklärung, kurz D-N-Erklärung, das Hempel in ([1942] 1965) entwickelte [...], zeichnet sich vorallem durch seine logische *Allgemeinheit* aus. (Schurz 2004, S. 6, Kursivierungen im Originalzitat)

Hempels ([1942] 1965) Hauptthese ist, dass die Geisteswissenschaften, speziell die Geschichtswissenschaften, auf die Suche nach allgemeinen Gesetzesmäßigkeiten »unvermeidlich angewiesen sind, sobald sie etwas zu erkennen suchen« (Schurz 2004, S. 158). In diesem Sinne entwickelt er das Modell der »deduktiv-nomologischen Erklärung« (Schurz 2004, S. 158):

Um auf möglichst viele Disziplinen anwendbar zu sein, verzichtet das Modell auf jedwede metaphysischen Annahmen über Erklärung im Sinne einer Wesensschau, sondern charakterisiert die Erklärung einer Tatsache E schlicht als deren logische Folgerung aus anderen Fakten und übergeordneten Gesetzeshypthesen, wobei Gesetzeshypthesen wiederum keine metaphysischen Kausalannahmen involvieren, sondern lediglich als strikte Regularitätsbehauptungen im Humeschen Sinn, sogenannte Allsätze der Form »Immer wenn A(x), dann B(x)«, verstanden werden.

Um diesen an den Naturwissenschaften orientierten Begriff der Erklärung als Ableitung des Einzelfalls vom allgemeinen Gesetz auf die Geistes- und Sozialwissenschaften anwenden zu können, bedarf es allerdings einiger Modifizierungen. Hempel selbst ([1942] 1965, S. 237) sieht in den Geschichts- und Sozialwissenschaften »kaum *strikte*, d. h. *ausnahmslos geltende* Gesetzesbeziehungen [...], statt dessen gäbe es »weiche« Gesetzesbeziehungen, die man als *statistische* Hypothesen auffassen könnte«. (Schurz 2004, S. 159, Kursivierungen im Originalzitat) Schurz

---

<sup>4</sup> Hier zitiert als »Hempel ([1942] 1965)«.



(2004, S. 160f.) führt einige Probleme eines starken »metaphysischen« Begriffs von ›Erklärung‹ an und resümiert, dass »man den Erklärungsbegriff besser nicht als rein *semantischen* Begriff einer zeitlich ›gültigen‹ Erklärung, sondern als einen auf ein gegebenes epistemisches Hintergrundsystem bezogenen und insofern *pragmatischen* Begriff expliziert.« (Kursivierungen im Originalzitat)

Ein Fallbeispiel zur Konkretisierung unterschiedlicher Begriffe von ›Erklärung‹ entnimmt Schurz (2004, S. 162) einem Text des Geschichtstheoretikers und Methodenpluralisten William Dray (1957, S. 33). Es bezieht sich auf Ludwig XIV, dessen Unpopularität in den Geschichtswissenschaften damit erklärt werde, »daß Ludwig der XIV sein Land oft in Kriege verwickelte und dem Volk schwere Belastungen aufbürdete.« Nach Dray müsste eine strenge, nomothetische Erklärung des Sachverhalts folgendermaßen lauten: »Alle Herrscher, die ihr Land in Kriege verwickeln und dem Volk schwere Belastungen aufbürden, werden unpopulär.« (zitiert nach Schurz 2004, S. 162). Schurz weist darauf hin, dass das ganz offensichtlich nicht stimmen könne. Es bliebe die in den Geschichtswissenschaften beliebte Normalfallhypothese oder normische Hypothese: »Herrscher bzw. Regierungen, die ihr Land in Kriege verwickeln und dem Volk schwere Belastungen aufbürden, werden *normalerweise* (üblicherweise, zumeist) unpopulär.« Nach Schurz (2004, S. 163) gleicht diese Art von Erklärung in ihrer untheoretisierten Form einer »Binsenwahrheit«. Sie ist offensichtlich nicht falsifizierbar, d.h. man kann nicht angeben, was der Fall sein muss, damit die Hypothese verworfen werden muss bzw. die Erklärung falsch ist. Schließlich seien Ausnahmen immer zugelassen und deren Ausmaß nicht weiter spezifizierbar. An späterer Stelle verweist Schurz (2004, S. 167), wiederum mit Verweis auf Hempel, auf die induktiv-statistische Erklärung als Lösungsansatz: »Die Wahrscheinlichkeit, daß ein Herrscher unpopulär wird, der sein Land in Kriege verwickelt, ist 84 %; Ludwig der XIV verwickelte sein Land in Kriege; also wird Ludwig der XIV mit bedingtem induktiven Wahrscheinlichkeitsgrad von 84 % unpopulär.« Auch wenn die induktiv-statistische Erklärung von Schurz (2004, S. 167f.) ebenfalls als unbefriedigend angesehen und mit Gegenargumenten bedacht wird (es wären z.B. viele zu erklärende Phänomene nicht quantifizierbar bzw. quantitativ beobachtbar), so wird zuerst in den Sozial- und nun auch in den Geisteswissenschaften eben die statistische als ›harte‹ Erklärung zunehmend populärer. Schurz selbst verweist auf die Kontextsensitivität historischer Sachverhalte, aus der deren Nicht-Verallgemeinerbarkeit folge, und präferiert eine modifizierte Form der normischen Hypothese, der er eine evolutionstheoretische Fundierung gibt:

Im Alltagsdenken sowie in allen wissenschaftlichen Disziplinen, von der Biologie aufwärts bis zu den Geisteswissenschaften, haben wir es überwiegend mit normischen Gesetzhypothesen zu tun: normalerweise können Vögel fliegen, handeln Menschen zweckrational, versuchen Regierungen die Wirtschaft ihres Landes intakt zu halten, funktionieren Lichtschalter, usw. – überall gibt es Ausnahmen. Was ist der Grund für diese Omnipräsenz normischer Gesetze – sind sie lediglich das Resultat einer subjektiven Zurechtinterpretation, Verschönerung einer in Wahrheit viel komplexeren Wirklichkeit, oder entspricht ihnen ein Realgrund, der normische Gesetze als genuinen Typ von nomologischen Gesetzesbeziehungen auszeichnet? (Schurz 2004, S. 170)

Sein Argument lautet, sehr verkürzt dargestellt, dass normische Erklärungen die Form nicht-numerischer induktiv-statistischer Erklärungen haben, ihnen also eine eindeutige – wenn auch komplexe und oft nicht quantitativ erfassbare – Realität zugrunde liege. Von deduktiven Schlüssen unterschieden sich normische Schlüsse durch ihre »Nichtmonotonie« (Schurz 2004, S. 171), d.h., dass eine Ableitung aus einem normischen, also mit Ausnahmen behafteten Gesetz so lange gilt, bis eine Ausnahme, die sich offensichtlich nicht unter der Ableitung subsummieren lässt, auftritt und auf ein spezifisches »Ausnahmegesetz« verweist, das die Erklärung des Normalfalls nicht außer Kraft setzt, sondern regelgeleitet in ihrer Geltung einschränkt. Als Beispiel gibt er Ableitungen über Eigenschaften eines Tiers, das wir als Vogel identifizieren:

Solange wir über dieses Tier nichts anderes wissen, als daß es ein Vogel ist, nehmen wir per default an, daß es sich um einen normalen Vogel handelt, und schließen, daß er fliegen kann. Dies ist ein wesentliches Prinzip der sogenannten nichtmonotonen Logik, die deshalb auch default logic genannt wurde: in Ermangelung gegenteiligen Wissens nehmen wir immer den Normalfall an. Sobald wir aber gegenteilige Evidenz erwerben, beispielsweise daß dieser Vogel einen gebrochenen Flügel hat, wird die Normalfallhypothese ›Vögel können normalerweise fliegen‹ durch das spezifischere normische Ausnahmegesetz ›Vögel mit gebrochenen Flügel können normalerweise nicht fliegen‹ blockiert: wir dürfen unseren ursprünglichen Schluß nicht mehr ziehen, sondern müssen das Ausnahmegesetz anwenden. Genau das ist Nichtmonotonie. (Schurz 2004, S. 171)

›Verstehen‹ ist in Schurz' einheitswissenschaftlichem Programm mit dem Erklären korreliert. Verstehen heißt ›etwas erklären können‹. (Schurz 2004, S. 169) Während diese Lösung des Verstehen-Erklären-Problems vielleicht etwas zu einfach und etwas zu naturwissenschaftlich gedacht ist, wenn es um Sprachverstehen und dessen heuristische Rolle in sprachbasierten akademischen Disziplinen geht, so helfen die von Schurz diskutierten verschiedenen Erklärungstypen, linguistische Programme des Erklärens voneinander unterscheiden zu können.

#### **4 Linguistik als erklärende Wissenschaft**

Die Linguistik ist seit ihrer Disziplinenwerdung in ihrem Kern als erklärende Wissenschaft verstanden worden. In der Tat zielt sie auf das Allgemeine, die Sprache. Als Datum dafür haben wir Linguistinnen und Linguisten aber nur das Individuelle, das Gesprochene und Geschriebene. Dass der Text oder das gesprochene Wort als Explikat einer allgemeinen Regel erst einmal verstanden werden muss, wurde und wird in vielen Bereichen der Linguistik als Alltagsgrundlage des Fachs aufgefasst, die außerhalb der Methodologie vorausgesetzt wird. Daher gibt es in den Standardtheorien der Linguistik auch keine Methodologien des regelgeleiteten fachlichen



**Tab. 1** Erklärungstypen in der Linguistik

Erklärungstyp	Kernbegriff	Beispiel
die nomothetische E.	Gesetz	Osthoff/Brugmann (1878–1910)
die strukturalistische E.	Relation im System	Saussure (1916)
die generative E.	Regel	Chomsky (1957)
die statistische E.	Muster	Labov (1972)
die funktionale E.	Kontext	Halliday (1973)
die praxeologische E.	Praxis	Gee (1992)
die diskurstheoretische E.	Formation	Reisigl/Wodak (2001)

Verstehens.<sup>5</sup> Wir können sechs Typen von linguistischen Erklärungen unterscheiden, die ich in Tabelle 1 aufführe, ohne ausführlich auf sie eingehen zu können. Dem Typ der Erklärung ist jeweils der explanative Kernbegriff und eine zentrale Publikation beigegeben (Tab. 1).

Einen großen Verwissenschaftlichungsschub in der Linguistik brachte das Programm der Junggrammatiker, das mit seinem Leitsatz von der Ausnahmslosigkeit der Lautgesetze den Prototyp einer an den Naturwissenschaften orientierten nomothetischen Erklärung ins Zentrum stellte. Die ›klassischen‹ Paradigmen des 20. Jhs., Strukturalismus und Generativismus, waren im wissenschaftstheoretischen Sinn jeweils an metaphysischen Gesetzesbegriffen orientiert, deren paradigmatische Geltung angenommen wurde, ohne dass sie experimentell oder durch Beobachtung falsifizierbar gewesen wären. Die statistische Erklärung ist aus der Soziologie über die Soziolinguistik in die Linguistik gekommen und hat angesichts einer immer größer werdenden quantifizierbaren Datenbasis zunehmenden Einfluss in nahezu allen Bereichen der empirischen Linguistik; kritische Stimmen z.B. zum Wert von Signifikanztests (Mulder 2020) bestätigen die zunehmende Bedeutung der Statistik eher. Rezente Erklärungsparadigmen wie das funktionale, das praxeologische und das diskurstheoretische weisen in unterschiedlicher Dringlichkeit auf die Bedeutungen von Kontexten, deren Variation und Nichtgeneralisierbarkeit für die linguistische Erklärung hin und entfalten unterschiedliche Fassungen dessen, was Schurz (2004, s. oben) die ›normische Erklärung‹ nennt. Damit konvergieren in vielen Fällen qualitative Methodologien, in denen das Verstehen, nämlich das Verstehen des Sprachdatums im Kontext eine zentrale Rolle spielt (Hermanns/Holly 2007; Felder/Mattfeldt 2015; Holly 2015). Da Forschungen in der empirischen Linguistik aller Spielarten aber – schon allein auf Grund der Datenverfügbarkeit und der technischen Entwicklung – immer häufiger auf digital repräsentierte, strukturell segmentierte und mit Metadaten ausgezeichnete Sprachkorpora zurückgreifen, gibt es einen gewissen Druck, einerseits die statistische Erklärung an irgendeiner Stelle miteinzubeziehen, schließlich sind Daten in dieser Form quantifizierbar und Hypothesen über deren Ausprägung und Distribution falsifizierbar. Andererseits gibt es aber das Bewusstsein, dass statistische Erklärungen eben (noch) keine linguistischen Erklärungen

<sup>5</sup> Was nicht bedeutet, dass es diese nicht gäbe. Am systematischsten ausgearbeitet liegt eine Grundlegung des linguistischen Textverstehens in Bärts (2015) 844 Seiten starken Kompendium zur hermeneutischen Linguistik vor.



**Tab. 2** Explikative Kaskade in der Korpuslinguistik

<b>Erklärungstyp</b>	Vorkategorisiertes Sprachdatum		statistische E.		strukturalistische E.		funktionale E.
<b>Kernbegriff</b>	Segment	→	Muster	→	Phrase	→	Konstruktion
<b>Beispiel</b> aus Hundt/ Oppliger (2022)	<i>But the fact was the pressure for a German economic presence came as much from the East Europeans as from the Germans.</i>		N-is		NP VA		Fokalisierungs- konstruktion

sind, da Sprache in Kontexten vollzogen wird und erst dann Sprache ist, wenn sie gemeint und verstanden wird (Müller 2012, 2015, S. 47–58). Insofern findet man in der zeitgenössischen empirischen Linguistik je spezifische Mischungen der genannten Erklärungstypen, bei denen z. B. der strukturalistische, der statistische und der funktionale Typ ineinandergreifen bzw. aufeinander aufbauen. In der digitalen Linguistik finden wir solche explikativen Kaskaden auf systematische Weise. Diese beginnen z. B. mit der statistischen Analyse und explizieren entsprechend erste Teilergebnisse als Muster, welche dann oft mit einer strukturalistischen Erklärung als sprachsystematisches Segment erklärt werden, z. B. als Nominalphrase. Diese können diskursfunktional erklärt werden, wie z. B. in der willkürlich, aber nicht zufällig ausgewählten Studie von Hundt/Oppliger (2022), in der eine Serie von Sprachdaten u. a. des Typs (*the*) *fact was / Tatsache ist ...* in basisannotierten Korpora als Instanzen eines statistischen Musters (*N-is*) erklärt werden, das wiederum in Begriffen der Phrasensyntax strukturalistisch beschrieben und schließlich diskursfunktional als Fokalisierungskonstruktion erklärt wird (Tab. 2).

## 5 Eine heuristische Nebenbei-Praxis: Verstehen in der digitalen Linguistik

“What we do when we analyse discourse using corpora ‘is a qualitative analysis of quantifiable patterns.” (Taylor/Marchi 2018, S. 6)

In dem hier vorangestellten Zitat aus einem Methodenbuch zur Diskursforschung mit Mitteln der Korpuslinguistik wird die in Tab. 2 dargestellte explikative Kaskade in vereinfachter Form angesprochen. Die angesprochene qualitative Analyse kann man auch unter den normischen Erklärungstypen subsummieren, die im Paradigma des Buchs von Taylor/Marchi (2018) funktional, praxeologisch und/oder diskurstheoretisch ausgedeutet werden. Praktisch geht es bei der qualitativen Analyse einerseits um Bildung oder Anwendung heuristischer Kategorien und andererseits um Kontextualisierung von Sprachdaten. Beides beinhaltet Verstehen. Das Zitat aus Taylor/Marchi (2018) steht programmatisch für die in dem Buch entfalteten Methodenvorschläge, die kenntnisreich und erkenntnistiftend Wege z. B. zum Umgang mit Abwesenheiten in Korpora (Duguid/Partington) und multimodalen Texten (Cagle) sowie Verfahren der Segmentierung (Marchi), Visualisierung (Anthony), Keyness-



Analyse (Gabrielatos) und reflexiven Forschung (Baker) aufzeigen, die alle selbstverständlich Verstehen voraussetzen. Nirgends wird das Verstehen selbst aber thematisiert, reflektiert oder methodologisiert. Diesen Befund kann man für die digitale Linguistik verallgemeinern. Eine Ausnahme bildet hier Diskussion um die Technik der Annotation, auf die ich im Folgenden eingehen möchte (s. unten).

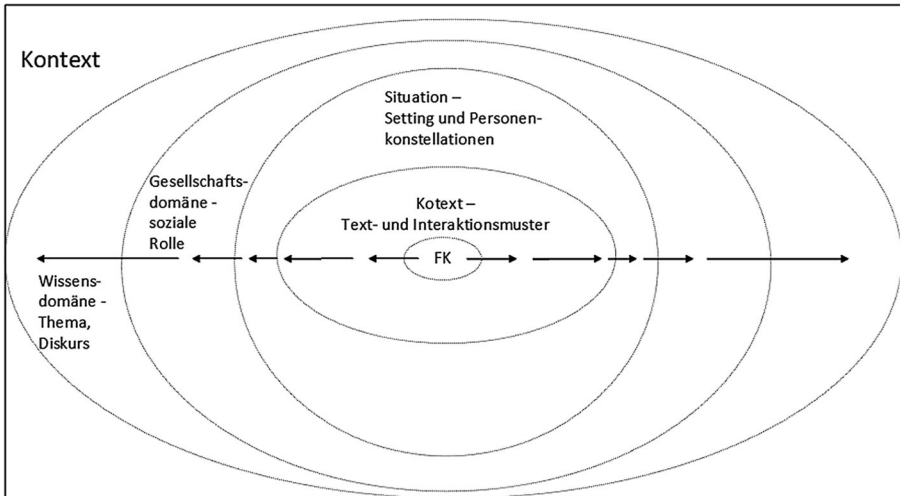
## 5.1 Was ist korpuslinguistisches Verstehen?

Zuvor möchte ich aber fragen, inwieweit die Verstehenstheorie, wie ich sie oben aus der Tradition der romantischen Hermeneutik heraus eingeführt habe, ein angemessener Rahmen für die digitale Linguistik sein kann. Dort geht es um Erkenntnis, die das Serielle, Typische, Regelmäßige in einer bestimmten Datenpopulation betrifft. Dass die einzelnen Sprachdaten Spuren von Interaktionen des Meinens und Verstehens sind (Müller 2012, 2015, S. 47–48), ist Voraussetzung jeder linguistischen Analyse. Es ist aber in aller Regel nicht im Mittelpunkt des Forschungsinteresses, was jemand in einer einzelnen Interaktion gemeint und verstanden hat.<sup>6</sup> Was interessiert, ist die Serialität der Interaktionen und die daraus erwachsenden Regularitäten, von denen wir aber umgekehrt annehmen können, dass sie die Voraussetzung für das individuelle Alltagsverstehen bilden. Der Zusammenhang von Mustererkennung und Verstehen ist auch in der neueren hermeneutischen Diskussion erkannt und thematisiert worden, besonders prägnant in der pragmatischen Verstehenstheorie Werner Kogges (2002):

Erstens sind sinnhafte Handlungen – mit Wittgenstein gesprochen – zwar durch relativ stabile Gepflogenheiten und Normalitäten des Gebrauchs geführt. Jedoch zwingt die Unterbestimmtheit der handlungsleitenden Muster gegenüber den je gegebenen Situationen zu einer ständigen, zumeist unauffälligen, da routinisierten Leistung des Verstehens. Diese Leistung besteht darin, in heterogenem und teilweise auch widerständigem Material ein unproblematisches Muster zu erkennen und herauszuschälen. Zweitens hat das Verstehen mit Wirkungen zu tun. Ob ich in einer Konfiguration diese oder jene Nuance betone bzw. als betont auffasse, kann sehr unterschiedliche Ketten von Konsequenzen nach sich ziehen. (Kogges 2002, S. 264)

Hier bildet die Fluchtlinie des Verstehens nicht mehr die Geisteswelt des meinentenden Individuums, sondern vielmehr die von den Interagierenden als gemeinsam vorausgesetzte Erfahrung mit der Materialität, Serialität und Kontextsensitivität sprachlicher Äußerungen. Das trifft die Hauptannahme der linguistischen Beschäftigung mit dem Verstehen. Hier ist der Kernbegriff »Kontextualisierung« (Gumperz 1982; Müller 2012, 2020). Dabei geht es darum, dass sprachliche Äußerungen neben der symbolischen immer auch eine indexikalische Bedeutung haben, die darauf verweist, mit welchen sprachlichen, situativen, sozialen und thematischen Kontexten man typischerweise zu rechnen hat, wenn etwas Bestimmtes gesagt oder geschrieben wird. »Kontextualisierung« ist also im Kern ein interaktionaler Begriff.

<sup>6</sup> Ausnahmen betreffen etwa die Schnittstellen der Korpuslinguistik zur Literaturwissenschaft und zur Gesprächsforschung.



**Abb. 1** Das Zwiebelmodell der Kontextualisierung – FK = Fokuskonstruktion. Aus: Müller (2012, S. 50)

Vorgeschlagen hat ihn John Gumperz (1982) im Rahmen der linguistischen Anthropologie. Kontextualisierung ist ein Verfahren, mit dem Äußerungen in einen von den Interaktanten selbst geschaffenen Kontext gestellt werden. Dieser Prozess verbindet empirische Beobachtungen mit Hintergrundwissen. Dieses wird in Form von Schemata organisiert und kann nach Auer (1986, S. 27) auf fünf verschiedenen Ebenen betrachtet werden: »Reden wir gerade miteinander?«, »Wer spricht gerade mit wem?«, »Was tun wir gerade?«, »Worüber sprechen wir gerade?« und »Wie stehen wir gerade zueinander?« Kontextualisierung hat in der empirischen Linguistik in verschiedenen Verwendungstraditionen an Bedeutung gewonnen. Der Begriff ist interaktional, epistemologisch und strukturell ausformuliert worden (Müller 2020, S. 46f.). Es wird jeweils mit unterschiedlichen Schwerpunkten die Bedeutung von indexikalischen Zeichenbeziehungen zwischen Äußerungen und ihrem Kontext für das Verstehen und die Bedeutungszuweisung betont. Die Kontextualisierungsfor schung hat mit diesem Modell gezeigt, dass es einen Zusammenhang zwischen der sprachlichen Musterbildung und der Typik situativer, sozialer und thematischer Kontexte gibt (Abb. 1). Damit bietet der Kontextualisierungsbegriff eine adäquate Heuristik, um Verhältnisse in Korpora zu erforschen, die über die Ausprägung von Mustern in Datensätzen hinausgehen, z. B. den Zusammenhang von Formulierungstradition und sozialer Rolle (Müller 2015) oder die sprachliche Markierung von Unsicherheit (Müller/Bartsch/Zinn 2021). Verstehen meint hier also die erfolgreiche Kontextualisierung eines sprachlichen Segments in einem gegebenen oder neu hergestellten Sinnzusammenhang. Es ist als Dimension alltäglicher sprachlicher Interaktion die Voraussetzung für die Erhebung und Analyse sprachlicher Datensätze (emische Kontextualisierung) und gleichzeitig als heuristisches Verfahren des regelgeleiteten Nachvollzugs von Prozessen der Sinnkonstitution ein zentrales Element der Forschungspraxis (etische Kontextualisierung – Müller 2015, S. 78f.).

## 5.2 Segmentierung und Verstehen

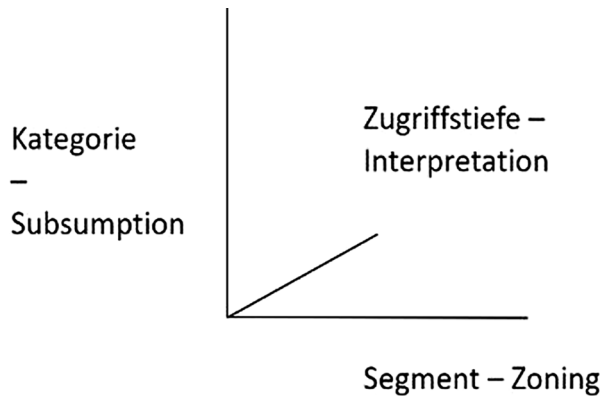
Man kann das Verstehen in der Korpuslinguistik also als regelgeleiteten Nachvollzug von Kontextualisierungsprozessen beschreiben. In korpuslinguistischen Studien wird es manifest meistens ganz am Ende, wenn Daten, die bereits statistisch, strukturalistisch und funktional erklärt wurden, diskutiert und in ihrer Bedeutung erfasst werden sollen – im Sinne von Taylor/Marchis (2018, S. 6) Aussage, die diesem Abschnitt als Zitat vorangestellt ist: das diskursanalytische Arbeiten mit Korpora sei eine qualitative Analyse quantifizierbarer Muster. Damit ist das Verstehen in der Korpuslinguistik aber nicht abgehandelt. Vielmehr spielt es als Nebenbei-Praxis eine allumfassende und ganz entscheidende Rolle in allen Schritten des korpuslinguistischen Verfahrens. Das kann man sich am Beispiel der Segmentierung klarmachen (Bartsch et al. 2023). Das Verfahren der Segmentierung von Wörtern, Phrasen, Sätzen oder Textabschnitten wird in der Korpuslinguistik eigentlich immer angewendet, teils als Hintergrundprozess in der Korpusanalyseumgebung. Segmentierung ist allerdings keineswegs eine rein formale Angelegenheit, sondern muss immer als hermeneutische Praktik im Sinne Kogges gedacht werden, in der drei Teilprozesse untrennbar ineinandergreifen und aufeinander bezogen sind (Abb. 2): Das Ausschneiden eines Segments aus dem Textkontinuum (*Zoning*), die Zuweisung des Segments zu einer analytischen Kategorie (*Subsumption*) und die Festlegung der Tiefe des gedanklichen Zugriffs auf der Basis von mehr oder weniger Kontextinformation (*Interpretation*).

Betrachtet man z.B. den Satz (a–1) isoliert und ohne Kontextwissen einzubeziehen, dann wäre er als deontisch modalisierte Proposition zu interpretieren und pragmatisch als direkter Sprechakt einzuordnen, konkret als Handlungsempfehlung. Berücksichtigt man aber den unmittelbaren Textzusammenhang (a–2), den Kontext, dann ergibt sich eine selbst gesteckte Zielsetzung als heuristische Textpraktik (Bender/Müller 2020) im Rahmen der Einleitung einer akademischen Arbeit, in Searles Terminologie ein kommissiver Sprechakt:

»a–1) *Die Arbeit soll in diesen [sic] Zusammenhang Aspekte darstellen, die bei der Erstellung eines solchen Verfahrens grundsätzlich zu beachten sind, und Wege aufzeigen, wie diese im konkreten Anwendungsfall zu einem anwendungsfähigen Verfahren konkretisiert werden können.*« (Dieleman 2016, S. 3)  
 a–2) *1.2 Zielsetzung der Arbeit. Ziel der Arbeit ist die Erarbeitung von allgemeingültigen Hinweisen für die Entwicklung von Entscheidungsverfahren, [...]. Die Arbeit soll in diesen [sic] Zusammenhang Aspekte darstellen, die bei der Erstellung eines solchen Verfahrens grundsätzlich zu beachten sind, und Wege aufzeigen, wie diese im konkreten Anwendungsfall zu einem anwendungsfähigen Verfahren konkretisiert werden können.* (Dieleman 2016, S. 3)

Ein anderes Beispiel betrifft die Segmentierung und Kategorisierung von Wörtern. Es stammt aus der als linguistisches Korpus aufbereiteten Version des Darmstädter Tagblatts (Stegmeier et al. 2022). Das Korpus wurde tokenisiert, also wortsegmentiert und mit dem für das Deutsche sehr oft verwendeten Stuttgart-Tübingen-Tagset (STTS – Schiller et al. 1999) wortartenannotiert. In Beleg b) sieht man, dass dabei zwangsläufig die diskontinuierlichen Partikelverben *abziehen*, *aufheben* und *ausru-*

**Abb. 2** Dimensionen der Segmentierung. Aus: Bartsch et al. (2023, S. 11)



*fen* als jeweils zwei Textwörter segmentiert werden. In der Kategorisierung wird zwar die Zugehörigkeit der Partikel zum Verbstamm ausgedrückt (PTKVZ = abgetrennter Verbzusatz), dennoch werden die Verben faktisch als diskontinuierliche Phrasen behandelt, z. B. bei der Wortzählung, die bei allen statistischen Operationen auf dem Korpus eine entscheidende Rolle spielt. Im Rahmen einer metaphysisch-strukturalistischen Erklärung würde man die Textwörter aber als Allomorphe von Derivations- und Stammmorphemen und Wortbestandteile klassifizieren und entsprechend 21 statt 24 Wörter zählen.

b-1) *Er zog hiebei seinen Hut ab, hob die Hände auf, und rief aus: Gott und dem Kaiser haben wir die guten Zeiten zu danken.* (Darmstädter Tagblatt, 01.03.1784)

b-2) *Er\_PPER zog\_VVFIN hiebei\_NE seinen\_NN Hut\_NN ab\_PTKVZ ,\_\$, hob\_VVFIN die\_ART Hände\_NN auf\_PTKVZ ,\_\$, und\_KON rief\_VVFIN aus\_PTKVZ :\_\$, Gott\_NN und\_KON dem\_ART Kaiser\_NN haben\_VAFIN wir\_PPER die\_ART guten\_ADJA Zeiten\_NN zu\_PTKZU danken\_VVINF .\_\$.*

Wenn man segmentiert, muss man also kategorisieren und interpretieren. Wichtig ist, dass man sich dessen bewusst ist und sich auch die Regeln, die man für die dem Segmentieren immanenten Verstehensprozesse geltend machen will, vor Augen führt. Umso mehr gilt das, wenn man mit Analyseumgebungen arbeitet, die Segmentierungsschritte automatisiert ausführen. Nach welchen Maßstäben erfolgt die Segmentierung und welche Interpretationsregeln sind dem Algorithmus eingeschrieben?

### 5.3 Annotation: Fokalisierungspraxis und Schnittstelle von Verstehen und Erklären

Eine sinnvolle Methode, Verstehensprozesse sich selbst und anderen explizit zu machen und außerdem für weitere, etwa statistische Verfahren verfügbar zu machen, ist die Annotation (Bender 2020; Bender/Müller 2020). Annotation ist eine Fokalisierungspraxis und bildet die Schnittstelle von Verstehen und Erklären. Dabei wird ein sprachliches Segment auf kategoriale Information bezogen. Man subsumiert

ein sprachliches Segment, z. B. ein Wort, eine Phrase oder einen Satz unter eine Kategorie, wie oben in Beleg b–2) für die Wortartenkategorisierung vorgeführt. Dazu muss man das Segment verstehen, also kontextualisieren. Da Verstehen, wie oben dargestellt, auf der Kommunikationsbiographie des Individuums beruht, ist natürlich mit subjektiven Interpretamenten beim Annotieren und daher mit unterschiedlichen Ergebnissen zu rechnen, wenn mehrere Personen denselben Datensatz annotieren. Das kann im Prozess interessant und hilfreich sein, weil es z. B. auf Polysemie und Vagheit in den Daten, auf nicht präzise genug zugeschnittene Annotationskategorien oder auf unzureichender Regelformulierung (Annotations-Guidelines) beruht. Wichtig ist aber erstens, solche Fälle zu identifizieren, und zweitens, am Ende einen annotierten Datensatz zu haben, der auf möglichst konsensualen Interpretationen beruht. Schließlich soll später beim Suchen oder Messen einer annotierten Kategorie sichergestellt sein, dass sich dahinter ein einheitlich beschriebenes Phänomen (z. B. Substantiv, Verbalphrase, direkter Sprechakt, Risikomarkierung, Argumentation) befindet. Ein wichtiges Merkmal der kollaborativen Annotation ist die flexible Herangehensweise an die Kategorienbildung. Kategorien können deduktiv aus theoretischen Ansätzen abgeleitet oder induktiv aus den Daten selbst entwickelt und differenziert werden. Dieser Ansatz erlaubt es, auch implizite semantische und pragmatische Phänomene zu erfassen, die eine tiefgehende Interpretation erfordern.

Insbesondere dann, wenn man manuelle Annotation zur Grundlage nehmen möchte, um einen Algorithmus zu trainieren, der die Annotation automatisch durchführt, ist es unerlässlich, dass die Segmentierungs-, Subsumptions- und Interpretationsverfahren – das Verstehen also – einheitlich gestaltet wurden (Becker/Bender/Müller 2020; Bender 2023; Bender/Becker/Kiemes/Müller 2023). Im Zentrum der kollaborativen Annotation steht also die gemeinsame Interpretation von Texten durch mehrere Annotierende. Eine Schlüsselkomponente ist die Erstellung eines Goldstandards, der als Referenz für die automatische Annotation dient. Um die Qualität der Annotationen zu gewährleisten, werden die Kategorien von verschiedenen Annotierenden überprüft, um Übereinstimmungen und Abweichungen zu identifizieren (Inter-Annotator-Agreement – Artstein 2017). Während kollaborative Verfahren der Annotation in der Computerlinguistik zum Zwecke der Automatisierung angewendet werden, kann man sie auch zur Kategorienentwicklung verwenden – zum Textverstehen in qualitativen Forschungssettings. Der Vorteil liegt dabei darin, dass die Annotation die verstehende Person dazu zwingt, Interpretamente zu externalisieren und genau zu sagen, auf welches sprachliche Segment sie es bezieht. Verstehen wird damit zu einem Verfahren, das in Gruppenprozessen angewendet werden und von Anderen in der Forschungsdebatte genau nachvollzogen diskutiert und – auf Basis eines in Guidelines und Metadaten angegebenen Kontext- und Regelwissens – auch falsifiziert werden kann. Damit ist die Annotation eine Schnittstelle zwischen pragmatischem Verstehen, normischer Erklärung und statistischer Analyse.

## 6 Anwendungsbeispiele

Die Vorteile und Herausforderungen des kollaborativen Verstehens möchte ich kurz an zwei Fallbeispielen verdeutlichen. Im ersten Fall geht es um die Analyse von



Praktiken des Moralisieren, zuerst in den Plenarprotokollen des Deutschen Bundestags, im zweiten um die kollaborative Analyse der Diskursgeschichte der Bundesrepublik Deutschland seit der Wiedervereinigung.

## 6.1 Praktiken des Moralisieren

Unter »Moralisierung« verstehen Felder/Müller (2022, S. 246)

die Einschreibung von moralischen Kategorien und Bewertungen in Debatten und den daraus erwachsenden diskursstrategischen Verweis auf eine Form der Unhintergebarkeit (Letztbegründung) eines Sachverhalts, der seine tatsächliche oder vermeintliche Gültigkeit dadurch erhält, dass er als moralischer Wert keiner weiteren Begründung oder Rechtfertigung bedarf.

Als Beispiel sei ein Beitrag Otto Schilys zur Rüstungsdebatte der frühen 1980er Jahre angeführt:

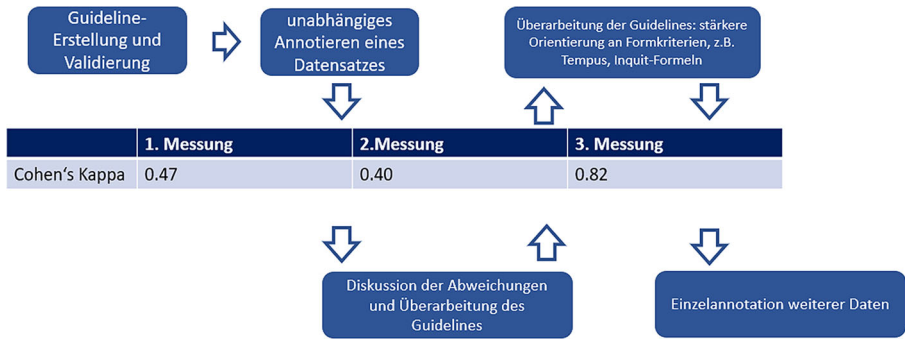
*Wenn man mit Massenvernichtungsmitteln droht – sie heißen zu Unrecht »Waffen« –, dann droht man mit einem vielfachen Auschwitz. (Otto Schily, Die Grünen, Rede im Deutschen Bundestag am 15.06.1983)*

Auf der Basis einer lexikalisch getriebenen Vorauswahl von Sätzen, die Moralvokabeln enthielten (Felder/Müller 2022, S. 249–254), haben die Autoren drei Stichproben von insgesamt 900 Sätzen aus den Jahren mit den meisten Moralvokabeln (1949, 1983, 2015) gezogen, doppelt annotiert und in drei Runden ein Inter-Annotator-Agreement gemessen (Abb. 3).

Die erste Annotationsrunde erbrachte ein zufriedenstellendes Ergebnis, wenn man bedenkt, dass Moralisierung ein vages, wertbehaftetes und weltanschaulich durchdrungenes Konzept ist [...]. Der statistische Agreement-Wert berücksichtigt, dass Übereinstimmungen ja auch zufällig erfolgen können (Artstein 2017). Die abweichenden Beispiele wurden dann diskutiert und die Guidelines entsprechend erweitert. In der zweiten Runde ist die Übereinstimmung dann allerdings nicht besser, sondern deutlich schlechter geworden. Es wurde deutlich, dass die Kriterien noch sehr stark von der Satzbedeutung und der gemutmaßten Sprecherintention ausformuliert waren und stark auf Kontexten basierten. Kontexte waren beim Annotieren aber gar nicht präsent, weil ja extrahierte Einzelsätze annotiert wurden. Wir haben daher die Guidelines ein weiteres Mal überarbeitet und diesmal auf Oberflächenindikatoren geachtet, formale Richtlinien erstellt und die Kategorie der Moralisierungspraktiken stark eingeschränkt. Auf diese Weise haben wir bei der dritten Messung ein Ergebnis erzielt, das auch in der computerlinguistischen Literatur als exzellente Übereinstimmung bezeichnet wird (vgl. Becker/Bender/Müller 2020, S. 442). (Felder/Müller 2022, S. 252)

Hier wurde also das Inter-Annotator-Agreement als Instrument zur Kategorienschärfung und Sicherung der gemeinsamen Verstehensbasis verwendet. Eine Auswertung ergab, dass der Anteil der Moralisierungen an den Moralthematisierungen in den Stichproben abnimmt. Während das Moralvokabular insgesamt in den Jahren





**Abb. 3** Kollaborative Annotation und Inter-Annotator-Agreement bei der Korpusanalyse von Moralisierungen

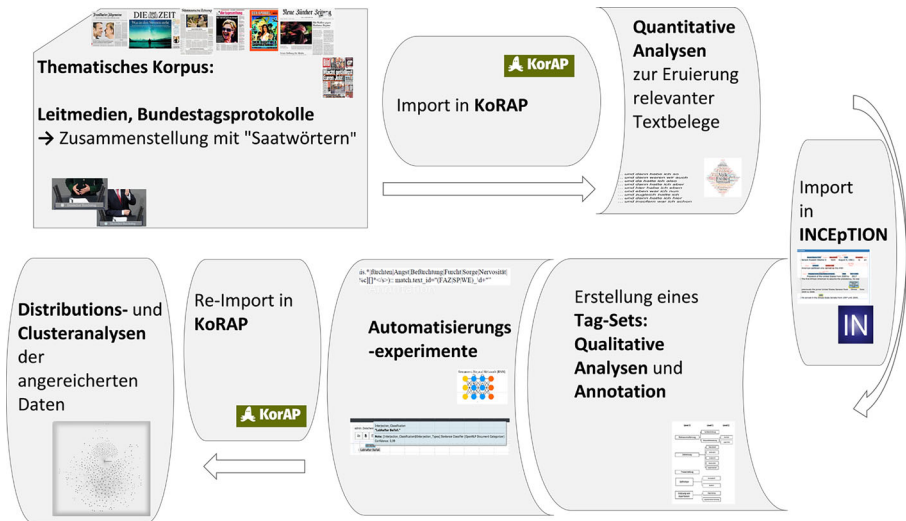
1983 und 2015 wesentlich höher ist, sind im Jahr 1949 anteilig Moralisierungen im eingeführten Sinne am häufigsten. Die Annotation von Moralisierungspraktiken wird auf dieser Basis unter Federführung der Computerlinguistin Maria Becker automatisiert und auf Daten aus weiteren Domänen angewandt (Becker/Felder/Müller 2023)

## 6.2 Gruppenverstehen: Kontroverse Diskurse

Ein methodisch ganz ähnliches Projekt, das aber die Methode skaliert im Hinblick auf Forschende und Kategorien ist die DFG-Forschungsgruppe *Kontroverse Diskurse. Sprachgeschichte als Zeitgeschichte seit 1990*, die im Juni 2022 ihre Arbeit aufgenommen hat (Wengeler 2022). Es geht dabei darum, die Sprachgeschichte seit der deutschen Wiedervereinigung als eine Geschichte der kontroversen sprachlichen Bearbeitung öffentlicher Themen zu rekonstruieren. Dabei soll Diskursgeschichtsschreibung zum ersten Mal in dieser Größenordnung als echte Gruppenforschung organisiert werden, was u. a. heißt, dass die Gruppe nicht nur eine gemeinsame Korpusinfrastruktur aufbaut und nutzt, sondern auch ein gemeinsames Annotationsschema entwickelt und damit Analysen einzelner Teilprojekte für die gesamte Gruppe fruchtbar macht. Dabei baut sie auf ein Prozessmodell, in dem algorithmische und interpretative Verfahren möglichst systematisch und transparent ineinandergreifen (Abb. 4).

Die Gruppe arbeitet in fünf Teilprojekten die zentralen Konzepte, Wissensbestände und Diskursstrategien der politischen und gesellschaftlichen Diskurse der letzten drei Jahrzehnte heraus. Vier Teilprojekte sind dabei inhaltlich nach sog. »tiefensemantischen Grundfiguren« (Partizipation & Egalität, Mensch & Technologie, Individuum & Gesellschaft, Freiheit & Sicherheit) organisiert, Teilprojekt 5 ist ein methodologisches. Es begleitet reflexiv den kollaborativen Annotationsprozess und führt Automatisierungsexperimente durch. Aus den verschiedenen Datenquellen (s. Abb. 4) stellt jedes Teilprojekt über geeignete Suchwörter (»Saatwörter«) ein Korpus themenrelevanter Texte zusammen. Dieses dient als Grundlage für Messungen, die sich Methodensets der Korpuslinguistik bedienen und das Ziel haben, geeignete Texte für dichte, interpretative Analysen zu eruieren. Dieses Kernkorpus werden in





**Abb. 4** Analyse-Workflow der DFG-Forschungsgruppe *Kontroverse Diskurse*. Aus: Müller (2023, S. 177)

die Annotationsplattform INCEPTION importiert und dort mit semantischen und pragmatischen Kategorien wie Metapher, Topos oder Präsupposition annotiert. Dazu erstellt die Forschungsgruppe inkrementell ein gemeinsames Annotationsschema. Insofern dabei reliable Tagger für die diskurssemantischen und -pragmatischen Kategorien entstehen, wird das Gesamtkorpus mit diesen Annotationen angereichert. Diese werden über die Korpusinfrastruktur durchsuchbar gemacht, so dass die Verteilung und Kombinatorik der qualitativ erhobenen Kategorien im Gesamtkorpus gemessen werden kann (zum Prozess Müller/Stegmeier 2022). Damit soll erreicht werden, dass erstens die Analysen der Teilprojekte möglichst eng verschränkt werden können und zweitens die Methoden der qualitativen, verstehenden Diskursgeschichte auf große Datenbestände möglichst ohne Qualitätsverlust angewendet werden können (Müller 2023, S. 177f.). Auch hier ist die Annotation also die zentrale Technik, in der pragmatisches Verstehen, statistische Analyse und normische Erklärung ineinandergreifen. Dazu ist die Gruppe so zusammengestellt, dass Expertinnen und Experten sowohl für hermeneutische als auch für statistisch-algorithmische Verfahren zusammenarbeiten und ihr Prozesswissen zusammenführen können.<sup>7</sup>

## 7 Fazit

In diesem Beitrag habe ich das methodenpluralistische Begriffspaar ›Verstehen‹ und ›Erklären‹ auf die Methodendiskussion der digitalen Linguistik angewendet. Wir haben gesehen, dass Verstehen im Sinne einer Ko-Konstruktion von Sinn eine

<sup>7</sup> Informationen über Zusammenstellung und Programm der Forschungsgruppe finden sich hier: [www.kontroverse-diskurse.net](http://www.kontroverse-diskurse.net) (15.09.2023).

grundlegende Voraussetzung für die linguistische Forschung ist. Das Verstehen von sprachlichen Äußerungen im Kontext ermöglicht die Analyse von Mustern, Regularitäten und Kontextualisierungsprozessen, die wiederum für das Erklären sprachlicher Phänomene bedeutsam sind. Mit Schurz (2004) habe ich verschiedene Typen des Erklärens unterschieden und diese dann auf die linguistische Tradition bezogen. Während Praktiken des Erklärens in der digitalen Linguistik gut reflektiert und methodologisiert sind, ist das Verstehen eine Nebenbei-Praxis, dessen Beherrschung als Alltagskunst vorausgesetzt und nicht weiter diskutiert wird. Verstehen und Erklären greifen in der digitalen Linguistik notwendigerweise ineinander. Ein besonderes Augenmerk verdient in diesem Kontext die Methode der Annotation, die als Schnittstelle von Verstehen und Erklären fungiert. Durch die kategoriale Zuordnung von sprachlichen Segmenten wird das Verstehen methodisch kontrolliert und für weitere Analysen verfügbar gemacht. In allen Bereichen der Linguistik sehen wir eine wachsende Bedeutung von statistischen Erklärungen, weil die quantifizierbare Datenbasis linguistischer Forschung immer weiter zunimmt. Dennoch darf man nicht vergessen, dass sprachliche Daten immer in einem Kontext interpretiert werden müssen, was das Verstehen als grundlegenden Schritt in der linguistischen Forschung unverzichtbar macht.

In der digitalen Linguistik ist Kontextualisierung als Dimension alltäglicher sprachlicher Interaktion die Voraussetzung für die Erhebung und Analyse sprachlicher Datensätze (emische Kontextualisierung) und gleichzeitig als heuristisches Verfahren des regelgeleiteten Nachvollzugs von Prozessen der Sinnkonstitution ein zentrales Element der Forschungspraxis (etische Kontextualisierung). Man kann das Verstehen in der Korpuslinguistik also als regelgeleiteten Nachvollzug von Kontextualisierungsprozessen beschreiben. Mit dem geläufigen Programm, welches das diskursanalytische Arbeiten mit Korpora als eine qualitative Analyse quantifizierbarer Muster beschreibt, ist das Verstehen in der Korpuslinguistik aber nicht abgehandelt. Vielmehr spielt es als Nebenbei-Praxis eine allumfassende und ganz entscheidende Rolle in allen Schritten des korpuslinguistischen Verfahrens.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.



## Literatur

- Apel, Karl-Otto (1955): Das Verstehen (eine Problemgeschichte als Begriffsgeschichte). In: Erich Rothacker (Hg.): *Archiv für Begriffsgeschichte: Bausteine zu einem historischen Wörterbuch der Philosophie*. Bd. 1. Bonn: H. Bouvier u. Co., S. 142–199.
- Artstein, Ron (2017): Inter-annotator agreement. In: Nancy Ide/James Pustejovsky (Hg.): *The Handbook of Linguistic Annotation – Introduction*. Dordrecht: Springer, S. 297–313.
- Auer, Peter (1986): Kontextualisierung. In: *Studium Linguistik* 19, S. 22–47.
- Bär, Jochen (2015): *Hermeneutische Linguistik. Theorie und Praxis grammatisch-semantischer Interpretation. Grundzüge einer Systematik des Verstehens*. Berlin/Boston: De Gruyter.
- Bartsch, Sabine/Gius, Evelyn/Müller, Marcus/Rapp, Andrea/Weitin, Thomas (2023): Sinn und Segment. Wie die digitale Analysepraxis unsere Begriffe schärft. In: *Zeitschrift für digitale Geisteswissenschaften* 8. [https://doi.org/10.17175/2023\\_003](https://doi.org/10.17175/2023_003).
- Becker, Maria/Bender, Michael/Müller, Marcus (2020): Classifying Heuristic Textual Practices in Academic Discourse: A Deep Learning Approach to Pragmatics. In: *International Journal of Corpus Linguistics* 25(4), S. 426–460.
- Becker, Maria/Felder, Ekkehard/Müller, Marcus (2023): Moral und Moralisierung. Linguistische Zugänge zu einem diskursrelevanten Phänomen. In: *Deutsche Sprache* 01, S. 26–50. <https://doi.org/10.37307/j.1868-775X.2023.01.04>.
- Bender, Michael (2020): Annotation als Methode der digitalen Diskurslinguistik. In: *Diskurse digital. Theorien – Methoden – Fallstudien* 2(1), S. 1–35. <https://doi.org/10.25521/diskurse-digital.2020.140>.
- Bender, Michael (2023): Pragmalinguistische Annotation und maschinelles Lernen. In: Lars Bülow/Konstanze Marx/Simon Meier-Vieracker/Robert Mroczynski (Hg.): *Digitale Pragmatik*. Stuttgart: Metzler, S. 267–286.
- Bender, Michael/Müller, Marcus (2020): Heuristische Textpraktiken. Eine kollaborative Annotationsstudie zum akademischen Diskurs. In: *Zeitschrift für Germanistische Linguistik (ZGL)* 48(2), S. 1–46. <https://doi.org/10.1515/zgl-2020-0001>.
- Bender, Michael/Becker, Maria/Kiemes, Carina/Müller, Marcus (2023): Category Development at the Interface of Interpretive Pragmalinguistic Annotation and Machine Learning – Annotation, Detection and Classification of linguistic routines of discourse referencing in political debates. In: *Digital Humanities Quarterly* 17(3) (Special Issue: Working on and with Categories for Text Analysis: Challenges and Findings from and for Digital Humanities Practices). In: <http://www.digitalhumanities.org/dhq/vol/17/3/000720/000720.html> (12.01.2024).
- Biere, Bernd Ulrich (1989): *Verständlich-Machen. Hermeneutische Tradition – historische Praxis – sprachtheoretische Begründung*. Tübingen: Niemeyer.
- Chomsky, Noam (1957): *Syntactic Structures*. Den Haag: Mouton.
- Dieleman, Owen (2016): *Hinweise für die Entwicklung von Verfahren zur maßnahmenübergreifenden Dringlichkeitsbewertung von Straßenbaumaßnahmen. Ein Beitrag zur Entscheidungsfindung im Rahmen der Aufstellung von Bauprogrammen für Straßenbaumaßnahmen*. Diss. Darmstadt. Online-Publikation: In: <http://tuprints.ulb.tu-darmstadt.de/5570> (12.01.2024).
- Dilthey, Wilhelm ([1894]1990): *Die geistige Welt*. Gesammelte Schriften V. Leipzig/Berlin: Teubner.
- Dray, William (1957): *Laws and Explanation in History*. London: Oxford Univ. Press.
- Droysen, Johann Gustav (1886): *Grundriss der Historik*. Leipzig: Veit. Online-Edition: Deutsches Textarchiv [https://www.deustchestextarchiv.de/droysen\\_historik\\_1868](https://www.deustchestextarchiv.de/droysen_historik_1868) (16.09.2023).
- Felder, Ekkehard/Anna Mattfeldt (2015): Linguistik als hermeneutische Wissenschaft. Das schwierige Verhältnis von Text und Bild im Diskurs. In: Reiner Keller/Werner Schneider/Willy Viehöver (Hg.): *Diskurs – Interpretation – Hermeneutik*. Weinheim: Beltz Juventa, S. 107–143.
- Felder, Ekkehard/Marcus Müller (2022): Diskurs korpuspragmatisch: Annotation, Kollaboration, Deutung am Beispiel von Praktiken des Moralisierens. In: Heidrun Kämper/Albrecht Plewnia (Hg.): *Sprache in Politik und Gesellschaft. Perspektiven und Zugänge (IDS Jahrbuch 2021)*. Berlin/Boston: De Gruyter, S. 241–261.
- Fetzer, James (2022): Carl Hempel. In: Edward N. Zalta/Uri Nodelman (Hg.): *Stanford Encyclopedia of Philosophy (Fall 2022 Edition)*. In: <https://plato.stanford.edu/archives/fall2022/entries/hempel/> (16.09.2023).
- Gadamer, Hans-Georg (1960): *Wahrheit und Methode: Grundzüge einer philosophischen Hermeneutik*. Tübingen: Mohr.
- Gee, James Paul (1992): *The social mind: Language, ideology, and social practice*. New York: Bergin & Garvey.



- Gumperz, John J. (1982): *Discourse strategies*. Cambridge, New York: Cambridge University Press.
- Halliday, Michael A. K. (1973): *Explorations in the Functions of Language*. London: Arnold.
- Hempel, Carl G. ([1942]1965): The Function of General Laws in History. In: Carl G. Hempel (Hg.): *Aspects of Scientific Explanation (and Other Essays)*. New York: Free Press, S. 231–243.
- Hermans, Fritz (2009): Verstehensdynamik. Zur Performanz des Verstehens. In: Angelika Linke/Helmuth Feilke (Hg.): *Oberfläche und Performanz. Untersuchungen zur Sprache als dynamischer Gestalt*. Tübingen: Niemeyer, S. 489–521.
- Hermans, Fritz/Werner Holly (Hg.) (2007): *Linguistische Hermeneutik. Theorie und Praxis des Verstehens und Interpretierens*. Tübingen: Niemeyer.
- Holly, Werner (2015): Diskurse verstehen? Optionen linguistischer Diskurshermeneutik. In: Reiner Keller/Werner Schneider/Willy Viehöver (Hg.): *Diskurs – Interpretation – Hermeneutik*. Weinheim: Beltz Juventa, S. 86–106.
- Hörmann, Hans (1976): *Meinen und Verstehen. Grundzüge einer psychologischen Semantik*. Frankfurt a.M.: Suhrkamp.
- Hundt, Marianne/Oppliger, Rahel (2022): (The) fact is ... /(Die) Tatsache ist ... focaliser constructions in English and German are similar but subject to different constraints. In: *International Journal of Corpus Linguistics* 27(1), S. 1–30. <https://doi.org/10.1075/ijcl.17073.hun>.
- Kogge, Werner (2002): *Die Grenzen des Verstehens. Kultur – Differenz – Diskretion*. Weilerswist: Velbrück Wissenschaft.
- Labov, William (1972): *Sociolinguistic Patterns*. Oxford: Blackwell.
- Liebert, Wolf-Andreas (2020): Hermeneutik und Empathie. In: Katharina Jacob/Klaus-Peter Koneiding/Wolf-Andreas Liebert (Hg.): *Sprache und Empathie. Beiträge zur Grundlegung eines linguistischen Forschungsprogramms*. Berlin/Boston: De Gruyter, S. 107–138.
- Mulder, Gerben (2020): The New Statistics for applied linguistics. In: *Dutch Journal of Applied Linguistics* 9(1-2), S. 79–96. <https://doi.org/10.1075/dujal.19019.mul>.
- Müller, Marcus (2012): Vom Wort zur Gesellschaft: Kontexte in Korpora: Ein Beitrag zur Methodologie der Korpuspragmatik. In: Ekkehard Felder/Marcus Müller/Friedemann Vogel (Hg.): *Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen*. Berlin/Boston: De Gruyter, S. 33–82.
- Müller, Marcus (2015): *Sprachliches Rollenverhalten. Korpuspragmatische Studien zu divergenten Kontextualisierungen in Mündlichkeit und Schriftlichkeit*. Berlin/Boston: De Gruyter.
- Müller, Marcus (2020): Kontextualisierung in der Re-Kontextualisierung. In: Simon Meier/Gabriel Viehauer/Patrick Sahle (Hg.): *Rekontextualisierung als Forschungsparadigma des Digitalen*. Norderstedt: IIDIE, S. 45–54.
- Müller, Marcus (2023): Corpus Approaches to Analysing Uncertainty and Ignorance in Academic Discourse. In: *Fachsprache* 45(1-2), S. 28–47. <https://doi.org/10.24989/fs.v45i1-2.2232>.
- Müller, Marcus/Stegmeier, Jörn (2022): Zur Methodologie der kollaborativen Diskursgeschichte. In: *Aptum, Zeitschrift für Sprachkritik und Sprachkultur* 18(3), Themenheft: *Kontroverse Diskurse*, S. 280–296.
- Müller, Marcus/Bartsch, Sabine/Zinn, Jens O. (2021): Communicating the unknown. An interdisciplinary annotation study of uncertainty in the coronavirus pandemic. In: *International Journal of Corpus Linguistics* 26(4), S. 498–531. <https://doi.org/10.1075/ijcl.21096.mul>.
- Osthoff, Hermann/Karl Brugmann (1878-1910): *Morphologische Untersuchungen auf dem Gebiete der indogermanischen Sprachen*. 6 Bde. Leipzig: Hirzel.
- Reisigl, Martin/Ruth Wodak (2001): *Discourse and Discrimination. Rhetorics of Racism and Antisemitism*. London et al.: Routledge.
- Saussure, Ferdinand de (1916): *Cours de linguistique générale*. C. Bailly/A. Secheyne (Hg.). Lausanne, Paris: Payot.
- Schiller, Anne/Teufel, Simone/Stockert, Christine/Thielen, Christine (1999): *Guidelines für das Tagging deutscher Textkorpora mit STTS. Technischer Bericht. Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart*. <https://www.ims.uni-stuttgart.de/documents/ressourcen/lexika/tagsets/stts-1999.pdf> (15.09.2023).
- Scholz, Oliver (2001): *Verstehen und Rationalität*. Frankfurt a.M.: Suhrkamp.
- Schurz, Gerhard (2004): Erklären und Verstehen: Tradition, Transformation und Aktualität einer klassischen Kontroverse. In: Friedrich Jaeger/Jürgen Straub (Hg.): *Handbuch der Kulturwissenschaften*. Bd. 2. Stuttgart: Metzler, S. 156–174.
- Stegmeier, Jörn/Günther, Christine/Hammer, Angela/Müller, Marcus/Stäcker, Thomas (2022): Eine Zeitung in drei Jahrhunderten: Digitalisierung des Darmstädter Tagblatts. In: *Information. Wissenschaft & Praxis* 73(2–3), S. 89–96.

- Taylor, Charlotte/Marchi, Anna (Hg.) (2018): *Corpus Approaches to Discourse: A Critical Review*. London: Routledge.
- Teichert, Dieter (2010): Erklären und Verstehen. Historische Kulturwissenschaften nach dem Methodenpluralismus. In: Jan Kusber/Mechthild Dreyer/Jörg Rogge/Andreas Hütig (Hg.): *Historische Kulturwissenschaften. Positionen, Praktiken und Perspektiven*. Bielefeld: transcript, S. 13–42.
- Wengeler, Martin (2022): Kontroverse Diskurse. Sprachgeschichte als Zeitgeschichte seit 1990. Ein Forschungsprojekt. In: *Aptum, Zeitschrift für Sprachkritik und Sprachkultur* 18(3), Themenheft: *Kontroverse Diskurse*, S. 201–206.
- Windelband, Wilhelm ([1894]1915): Geschichte und Naturwissenschaft. In: Wilhelm Windelband (Hg.): *Präludien: Aufsätze und Reden zur Philosophie und ihrer Geschichte*. Bd. 2. Tübingen: Mohr. <https://doi.org/10.11588/diglit.19222>.

**Hinweis des Verlags** Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.