

# Marc Kupietz und Thomas Schmidt

## Einleitung

Die Rolle korpusbasierter und korpusgestützter Sprachforschung hat in den letzten 15 Jahren einen Wandel durchlaufen wie wohl kein anderer Bereich der Linguistik. War Korpuslinguistik noch am Anfang der Nullerjahre in Deutschland vornehmlich im Bereich der diachronen Forschung, der Variationslinguistik und der Lexikologie relevant und darüber hinaus teilweise stark von der Computerlinguistik vereinnahmt, so ist sie heute im Mainstream fast aller linguistischer Subdisziplinen fest etabliert. Erfreulicherweise hat sich dabei auch gezeigt, dass qualitative und quantitative linguistische Ansätze sich keineswegs widersprechen, sondern dass gerade ihre Kombination den größten Erkenntnisgewinn verspricht.

Dass das IDS diesen Wandel nicht nur (auch) erfährt, sondern aktiv mitgestaltet hat, zeigt sich unter anderem in der Einrichtung der Programmbereiche „Korpuslinguistik“ (2004) und „Mündliche Korpora“ (2016) und mehrerer Großprojekte in den Abteilungen, etwa „Korpusgrammatik“ (Abteilung Grammatik), „Empirische Methoden“ (Abteilung Lexik) oder „Variation des gesprochenen Deutsch“ (Abteilung Pragmatik), für die korpuslinguistische Zugänge zu Sprache jeweils zentral sind. Nicht zuletzt hat auch der Ausbau der schriftlichen Korpora (Deutsches Referenzkorpus DeReKo) und der mündlichen Korpora (Forschungs- und Lehrkorpus Gesprochenes Deutsch FOLK, Korpora im Archiv für Gesprochenes Deutsch) als Daueraufgabe des Instituts an Bedeutung gewonnen, was sich auch in der Weiterentwicklung der zugehörigen Korpus-Plattformen (KorAP als Nachfolger von COSMAS I/II und die DGD2 als Nachfolger von DIDA und der DGD des DSAV) niederschlägt. Die jeweils 1.000 bis 2.000 neuen Nutzer, die sich jährlich für diese Plattformen registrieren, belegen deutlich die große Verbreitung korpuslinguistischer Methoden in der sprachwissenschaftlichen Forschung und Lehre über das IDS hinaus.

Entsprechend der allgemeinen Etablierung empirischer Verankerung in Korpusdaten und der Verwendung korpuslinguistischer Methodik kann die Aufgabe, einen Überblick über das Themenfeld „Korpuslinguistik“ zu verschaffen, wenn überhaupt nur von der gesamten Reihe geleistet werden. In diesem Band konzentrieren wir uns daher – im Sinne einer Bestandsaufnahme – auf die Grundlagen der Korpuslinguistik (Mair, Kehrein & Vorberger, Dipper & Kwekkeboom, Boas & Fingerhuth, Draxler & Schiel, Schmidt), auf einige interessante neuere Anwendungsfelder (Scharloth, Gries, Imo & Weidner), sowie auf Methodik (Bubenhofer, Scharloth) und Werkzeuge (Dipper & Kwekkeboom, Kehrein & Vorberger, Draxler & Schiel, Schmidt) zur Gewinnung linguistischer Hypothesen und Erschließung linguistischen Wissens aus Korpora.

Den Anfang macht Christian **Mair** mit einem Überblick über eine fünfzig-jährige *Erfolgsgeschichte Korpuslinguistik?*, der einerseits die zunehmende Verfügbarkeit von Korpusressourcen lobt, andererseits aber auch eine stärkere Einbeziehung gesprochener Spontansprache, den Aufbau mehrsprachiger Korpusressourcen und für die Korpuslinguistik insgesamt eine stärkere Integration mit den Digital Humanities anmahnt und für die Zukunft empfiehlt.

Noah **Bubenhofer** klassifiziert in seinem Beitrag *Visualisierungen in der Korpuslinguistik* diese zunächst nach verschiedenen Grundfiguren und Anwendungszielen, wirbt für das das mittelbare Erkenntnispotential gerade explorativer Visualisierungen als Zwischenschritt des Forschungsprozesses und diskutiert entsprechende Methodologien bzw. Workflows am Beispiel des Einsatzes von Kartendiagrammen zur Exploration von Geokollokationen und anhand der Analyse narrativer Muster in Geburtsberichten mithilfe unterschiedlicher Visualisierungstypen.

Joachim **Scharloths** Beitrag *Korpuslinguistik für sozial- und kulturanalytische Fragestellungen* wirbt für die Übertragung von Prinzipien der *Grounded Theory* auf korpuslinguistische Forschungsprozesse und demonstriert, wie man mit einer entsprechenden datengeleitete korpuslinguistischen Methodik die Analyse von Bundespressekonferenzen zu einem Modell der kommunikativen Gattung führen kann, das zum Verständnis ihrer gesellschaftlichen Funktion und damit zu sozial- und kulturwissenschaftlichem Erkenntnisgewinn beiträgt.

Auch Stefan Th. **Gries** Beitrag *Korpuslinguistik und ihr Potenzial für die (amerikanische) Rechtsprechung* erweitert das Anwendungsfeld korpuslinguistischer Methodik. Er zeigt anhand von zwei bekannten, linguistisch aber zweifelhaften Entscheidungen des Obersten Gerichtshofs der Vereinigten Staaten von Amerika, wie sich mithilfe einer methodisch fundierten Analyse von Konkordanzen und Kollokaten, der im amerikanischen Rechtssystem wichtige und weit verbreitete, aber bisher schlecht definierte Begriff des *ordinary meaning* besser fassen ließe und sich so die amerikanische Rechtsprechung durch die Anwendung korpuslinguistischen Wissens entscheidend verbessern ließe.

Stefanie **Dipper** und Sarah **Kwekkeboom** beschreiben im ersten Teil ihres Beitrags *Historische Linguistik 2.0* die Herausforderungen, die beim Aufbau des Verbundes Historischer Referenzkorpora des Deutschen zutage getreten sind. Im zweiten Teil demonstrieren und diskutieren sie die vielfältigen Möglichkeiten, das online frei verfügbare und detailliert dokumentierte Referenzkorpus Mittelhochdeutsch (ReM) mithilfe von ANNIS im Rahmen historisch-linguistischer Forschung zu nutzen.

Die Bestandsaufnahme zu verschiedenen Typen mündlicher Korpora eröffnen Roland **Kehrein** und Lars **Vorberger** mit ihrem Beitrag zu *Dialekt- und Variationskorpora*. Deren Tradition lässt sich weit in die vordigitale Zeit zurückführen, wo die Dialektforschung bereits im 19. Jahrhundert als eine der ersten

sprachwissenschaftlichen Disziplinen systematische Datenerhebungen zur Erstellung von Dialektatlanten durchgeführt hat. Mit den von Eberhard Zwirner in den 1950er Jahren auf Tonband erhobenen Daten zu den deutschen Mundarten entstanden dann die ersten mit dem Begriff „Korpus“ bezeichneten Sammlungen, die für die folgenden Jahrzehnte vorbildhaft waren und ihre moderne Entsprechung in Korpora finden, die auf regionalsprachliche statt „klassisch“ dialektale Variation fokussieren. Der Beitrag schließt mit einem Überblick über digitale Instrumente zum Zugriff auf diesen Typus von Korpusdaten.

Hans C. **Boas** und Matthias **Fingerhuth** widmen sich in ihrem Beitrag zu *Deutschen Sprachinselnkorpora im 21. Jahrhundert* dann der aktuellen Situation der Dokumentation deutsch(-basiert)er Varietäten außerhalb des deutschsprachigen Kerngebiets. Nach einem historischen Abriss des Forschungsgebiets geben die Autoren einen Überblick über aktuell an verschiedenen Standorten archivierte Sprachinselnkorpora und argumentieren für eine stärkere Integration und Koordination dieser verteilten Ressourcen, nicht zuletzt um perspektivisch deren korpuslinguistisches Potenzial besser nutzbar zu machen.

Der Beitrag von Christoph **Draxler** und Florian **Schiel** befasst sich mit der Erstellung und Datenaufbereitung von bzw. für *Moderne Phonetische Datenbanken*. Wie die Autoren zeigen, stehen solche Korpora typischerweise (und in Analogie zur Wechselwirkung zwischen schriftsprachlichen Korpora und Computerlinguistik) an einer Schnittstelle zwischen phonetischer Analyse und angewandter Sprachtechnologie. Es werden zunächst die betreffenden Datentypen beschrieben, bevor anhand eines Beispiels Prozesse der Erstellung und Nutzung eines phonetischen Korpus illustriert werden.

Die Bestandsaufnahme zu mündlichen Korpora beschließt Thomas **Schmidt** mit einem Beitrag zu *Gesprächskorpora*. Nach einer Abgrenzung dieses Korpus-typs gegenüber anderen (mündlichen) Korpora werden die wichtigsten seit den 1970er-Jahren erhobenen Gesprächskorpora vorgestellt. Das aktuell im Aufbau befindliche Forschungs- und Lehrkorpus Gesprochenes Deutsch (FOLK) bildet dann den Ausgangspunkt zu einer Diskussion aktueller und künftiger Herausforderungen, die methodologische Innovationen an der Schnittstelle zwischen Korpuslinguistik und Gesprächsforschung notwendig erscheinen lassen.

Im abschließenden Beitrag des Bandes nehmen Wolfgang **Imo** und Beate **Weidner** mit ihrer Diskussion von Einsatzmöglichkeiten *Mündlicher Korpora im DaF- und DaZ-Unterricht* dann eine Anwendungsperspektive ein. Dabei gehen sie zunächst auf die Rolle „authentischer“ gesprochener Sprache im Fremdsprachenunterricht ein und stellen mündliche Korpora vor, die für Unterrichtszwecke zur Verfügung stehen. Anschließend wird mit der Plattform Gesprochenes Deutsch eine Datensammlung vorgestellt, die speziell an den Bedürfnissen von Lehrenden und Lernenden in den Bereichen DaF und DaZ ausgerichtet ist.

