

Rezension zu

Noah Bubenhofer (2009): *Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*. Berlin/New York: Walter de Gruyter. (= *Sprache und Wissen* 4).

Sibylle Reichel (Bern)

Die Diskursanalyse, wie sie in jüngerer Zeit ins Zentrum der Forschung gerückt ist, liefert den Ausgangspunkt für "Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse", Noah Bubenhofers Dissertation an der Universität Zürich. Sein Interesse liegt darin, mittels einer linguistischen Oberflächenanalyse einen empirischen Beitrag zur Diskursanalyse vorzustellen. Neu an dem Ansatz ist die induktive Vorgehensweise, unvoreingenommen bzw. "naiv" an sehr große Datenmengen heranzugehen und diese in Hinblick auf wiederkehrende Muster im Sprachgebrauch zu analysieren. Eine solch umfangreiche Analyse ist nicht ohne Computereinsatz zu bewerkstelligen und so enthält die Arbeit auch eine ausführliche Einführung in die Korpuslinguistik nebst statistischen und computerlinguistischen Ausführungen. Die im ersten Teil entwickelte Methode wird anhand einiger Beispielanalysen über ein Zeitungskorpus der "Neuen Zürcher Zeitung" demonstriert.

Das Werk ist in vier größere Abschnitte gegliedert, die aus Theorie (I), Herleitung (II), Anwendung (III) und Fazit (IV) bestehen. Zu Beginn des Buches, im eigentlichen Theorieteil, definiert Bubenhofer, was er unter Sprachgebrauchsmustern versteht. Begrifflichkeiten wie *Schema*, *Schablone*, *Regel* und *Modell* werden präzisiert (29), um schließlich *Muster* bzw. *Sprachgebrauchsmuster* als Terminus technicus zu definieren. Als Erscheinungen, die sich an der Textoberfläche zeigen und die als "charakteristische Eckpfeiler von Diskursen gelesen werden [können]", bieten Sprachgebrauchsmuster einen lohnenden Ansatzpunkt für die Diskurs- und Kulturanalyse. Leider weist dieser terminologische Teil einige Redundanzen auf, während man an dieser Stelle eine im einleitenden Kapitel sehr knapp und erst im zweiten Abschnitt des Buches ausführlicher vorgenommene Definition der Termini *Kollokation* bzw. *Kookkurrenz* vermisst.

Es folgt ein Kapitel über Diskurse (31–39), das vom Diskursbegriff nach Foucault (1970 und 1981) ausgeht und zwei weitere Ansätze (Busse/Teubert 1994 und Warnke 2002) kontrastiert. Als Fazit wird der Einsatz von korpuslinguistischen Methoden in Form einer quantitativen Untersuchung von Sprachgebrauchsmustern mit einer linguistischen Analyse quer über verschiedene Texte vorgeschlagen. Dies stellt laut Bubenhofer einen direkteren Anschluss an die Foucault'sche Diskursanalyse dar als die kritisierte Analyse thematisch zusammengestellter Texte: "Klar ist jedoch, dass ein nach dem Kriterium des typischen Sprachgebrauchs definierter Diskurs zu anderen Analysen führt, [sic] als thematisch ausgerichtete Diskursanalysen" (37). Das Vorgehen ist operationalisierbar, da eine induktive Auswahl nicht nur des Korpus, sondern auch der Analyseeinheiten stattfindet und die Aufmerksamkeit auf Aussagen und nicht auf Texte gelenkt wird (38). In einem sehr knappen Exkurs bezieht der Autor auch die Kulturwissenschaft unter ethnologischen, historischen und linguistischen Gesichtspunkten in seine Überlegungen mit ein.

Der folgende letzte Teil der Theorie beleuchtet Muster im Sprachgebrauch aus einer soziolinguistischen Perspektive und leitet dann über zu einem eher semantischen Ansatz mit sechs Analysekategorien, welche einen Zugang zum Diskurs herstellen sollen. Hier wird die Arbeit anwendungsorientierter. Etwas befremdlich wirkt zunächst die Reihung der inhomogenen Termini *Stil*, *kommunikative Gattung*, *Mentalitäten*, *Typik des Verhaltens/Idiomatik*, *Argumentationsfiguren/Topoi* und *Metaphern* (55–83), bevor diese Konzepte dann erklärt und anschaulich miteinander in Beziehung gesetzt werden. Der Vergleich der verschiedenen semantisch-pragmatischen Phänomene und Theorieansätze findet mit einer bemerkenswerten Technik statt: Sogenannte Spinnenprofile machen die Konzepte vergleichbar (90f.). Eine grafische Darstellung, wie man sie heute (besonders aus der Schweizer) politischen Berichterstattung kennt (cf. Hermann/Leuthold 2008), stellen so die Forschungsansätze auf eindrück-

liche Weise einander gegenüber. Im Unterschied zu den ursprünglich für die Soziologie entwickelten und heute in den Sozialwissenschaften gebräuchlichen Netzdiagrammen (cf. Mayr 1877: 78) sind die Achsen bei Bubenhofer komplementär verteilt (z. B. *oberflächlich* vs. *tiefensemantisch*, *induktiv* vs. *deduktiv*) und gewähren so einen geordneten visuellen Eindruck. Schließlich macht der Autor damit die Stärken des korpuslinguistischen Ansatzes deutlich und kündigt für das Folgende eine Einführung in korpuslinguistische Methoden an.

Im zweiten großen Abschnitt geht es um die Entwicklung einer Methode zur Diskursanalyse mittels Mustererkennung dar (II Mustererkennung als Basis für die Analyse: Methodische Herleitung). An dieser Stelle werden nun Formen von Mustern, welche an der Textoberfläche zu finden sind, mit linguistischen Mitteln beschrieben. Bubenhofer definiert die Termini *Ko-Okkurrenz* und *Kollokation* und führt aus, dass ein erweitertes Verständnis von Kookkurrenz in Syntagmatischen Mustern mündet. Eine weitere Argumentationsstruktur in diesem Kapitel kommt zum Schluss (127), dass die Arbeit mit nicht-annotierten Korpora sinnvoll sei, da eine Annotation und damit eine Lemmatisierung einem Informationsverlust gleichkomme. Ein bisschen schnell wird über die Möglichkeit hinweggegangen ein nicht nur Wortarten-Lemma-annotiertes Korpus zu verwenden, sondern ein morphologisch oder gar semantisch annotiertes; das ist insofern berechtigt, als eine solche Annotation ja nicht unproblematisch ist und sich auch nicht so einfach automatisieren lässt - einen Hinweis darauf vermisst man aber an dieser Stelle. Man findet ihn wiederum als Desiderat ganz am Ende der Arbeit (335): Im Fazit wird das Problem der Annotation thematisiert und überdies darauf hingewiesen, dass auch eine semantische Annotation "Vorannahmen über potenzielle diskursemantische Grundfiguren erfordern würde" (314), welche der Autor vermeiden möchte.

Die Arbeit fasst die Korpuslinguistik grundsätzlich nicht als Instrument für die Linguistik, sondern als eigenständige Disziplin auf. Nach Klärung der Vorannahmen erfolgt nun eine statistisch-mathematische Herleitung der computerlinguistischen Methode. Verständlich, sauber und nachvollziehbar kommt der Autor zum Ergebnis, in erster Linie einen Log-Likelihood-Test und den χ^2 Test für die Filterung statistisch auffälliger Kookkurrenzen zu verwenden (146). In diesem korpuslinguistisch orientierten Abschnitt werden die statistischen Berechnungsalgorithmen gegeneinander abgewogen und diejenigen ausgewählt, die in der späteren konkreten Analyse zum Einsatz kommen. Die Einführung in die Korpuslinguistik und die Ausarbeitung der Auswertungswerkzeuge stellt den Kern der Arbeit dar und bietet eine umfassende Demonstration dessen, wie automatische Korpusanalysen strukturiert und entworfen werden. In diesem Zusammenhang liegt dem Autor daran, den Unterschied von corpus-based zu corpus-driven deutlich zu machen und anhand konkreter Beispiele zu erklären (149–174). Es wird betont, dass die entwickelte Methode einen induktiven (corpus-driven) Weg geht und nicht, wie andere Diskursanalysen einen deduktiven (corpus-based).

Zum Abschluss des Methodologieteils finden verschiedene Werkzeuge und Ressourcen für die computerlinguistische Auswertung Erwähnung, indem einige elektronische Korpora als die "wichtigsten öffentlich verfügbaren deutschsprachigen Korpora geschriebener Sprache" (175) aufgezeigt werden. Dabei werden nur die größten Korpusansammlungen berücksichtigt und im übrigen auf andere Quellen verwiesen (175). Ebenso wird die Auswahl der Software zur Kollokationszählung kurz wiedergegeben (182).

Schließlich erfolgt eine Demonstration der Analyse anhand mehrerer Beispiele im Teil III: Anwendungsbeispiele (187–304). Das im Teil II erarbeitete Konzept wird in Form von drei Einzelanalysen unterschiedlicher Ressorts aus einem Zeitungskorpus der NZZ angewandt. Das Gesamtkorpus "umfasst eine Zufallsauswahl von 44843 Artikeln aus dem Zeitraum von 1995 bis 2005" (189). Die generelle Vorgehensweise wird nachvollziehbar erklärt. Auch technische Details von Problemen mit inkonsequenter Umlautschreibung (196) bis hin zu Grenzen der Rechnerkapazität (197) werden erläutert (193–208).

Die erste konkrete Analyse besteht aus einem Vergleich zweier Zeiträume (1995–97 und 2003–05) aus dem Ressort *Auslandsberichte*. Zunächst werden Mehrworteinheiten computativ ermittelt, die dann unter Einbeziehung von Weltwissen für die eigentliche Analyse gefiltert werden. Der Autor bestätigt mit seiner korpuslinguistischen Methode unter anderem den nicht überraschenden Schluss, dass es "bei den zählenswerten Entitäten zeitgebundene Präferenzen für bestimmte sprachliche Realisierungen zu geben [scheint]" (240). Die Auswertung der *Inlandsberichterstattung* über dieselben beiden Zeiträume demonstriert das Potential der Methode, verschiedene Untersuchungsinteressen zu bedienen. So wäre es

möglicherweise auch denkbar, auf diesem Wege nationale Regionalismen zu identifizieren (245). Allerdings konzentriert sich der Autor auf das Thema seiner Arbeit, die Diskursanalyse. Er stellt unter anderem fest, dass "[d]ie Korpusanalyse [...] auf ein Sprachgebrauchsmuster (NICHT NUR ... SONDERN) aufmerksam gemacht [hat], das per se die Tendenz aufweist, bestimmte pragmatische Funktionen wie Argumentation anzuzeigen" und fragt: "Was bedeutet es, wenn ein Thema X plötzlich stärker in argumentativen Kontexten vorkommt?" Die dritte Beispielanalyse fasst die Resorts *Feuilleton*, *Sport* und *Wirtschaft* zusammen. Zur Darstellung von geographischen Bezeichnungen kommen bei dieser Analyse auch Karten zum Einsatz. Unter Einbeziehung seines Weltwissens äußert der Autor schließlich die Vermutung, dass "vermehrte Ethnienbezeichnungen auf kriegerische Konflikte schließen lassen" (280). Er schränkt aber ein, dass die Methode zu vereinfachend ist, um diese Aussage empirisch zu stützen. Weitere Einzelbeobachtungen beschäftigen sich etwa mit der Verwendung von *Swiss Ski* im Vergleich zu *Schweizer Skiverband* und dem Ersatz des Lexems *Damen* durch *Frauen* in der Sportberichterstattung.

Sehr vorsichtig formuliert Bubenhofer aufgrund der drei Beispielanalysen vier Hypothesen, deren Interpretation einen Beitrag zur genauen Beschreibung von Diskursen liefern könnten. Sie beziehen sich auf themengebundenes (14.1 Sprechen über KRIEG UND GEWALT) und kontextabhängiges (14.2 Sprechen über SORGEN UND PROBLEME) sowie zeitlich gebundenes (14.3 Sprachgebrauch und Welt) und textsortenabhängiges (14.4 Sprachgebrauch und Text) Auftreten von Mustern im Sprachgebrauch.

Der abschließende vierte Teil der Arbeit (IV Fazit und Ausblick) enthält eine ausführliche Evaluation der Methodik in Hinblick auf die Fragestellungen. Im Ausblick, der eine weitere Kurzanalyse eines weiteren Probekorpus aus Forumstexten einer online verfügbaren Politikplattform enthält, ermutigt der Autor dazu, weitere Korpora mit der entwickelten Methode zu analysieren.

Diskurse im Hinblick auf Sprachgebrauchsmuster zu untersuchen und zu beschreiben, scheint ein lohnender Ansatzpunkt zu sein, der durch die entwickelte Methode konkret nachvollziehbar gemacht wird. Eine Kritik, die oft bei Korpusanalysen aufkommt, dass die Herkunft der Daten oft unklar bleibt und man bei auffälligen Phänomenen auf Besonderheiten des jeweiligen Korpus stößt, ist dem Autor durchaus bewusst. Er berücksichtigt, dass immer nur Aussagen über ein bestimmtes Korpus möglich sind (60), und sich die Frage stellt, ob dies ausreichend für eine Diskursanalyse im Sinne Foucaults ist. Dennoch wird begründet extrapoliert und in den Beispielanalysen zeigen sich so auch interessante Schlüsse.

Bei den Erläuterungen zu den Musterdefinitionen am Anfang in der Einleitung hätte eine Konsolidierung der bisweilen in den Gedankengängen sprunghaften Argumentationsstruktur den Ausführungen womöglich gut getan. Demgegenüber steht der knappe, gestraffte Abriss über Diskurs- und der noch kürzere über die Kulturanalyse. Wer in diesem Teil eine komplette Einführung in die Diskursanalyse erwartet, der wird enttäuscht. So knapp die Einführung in die Diskurs- und Kulturanalyse ist, so ausführlich und einleuchtend ist die Einführung in die Korpusanalyse und die computerlinguistischen sowie statistischen Ansätze. Das Werk ist somit vor allem für Personen interessant, die aus der Diskursanalyse kommen und die eine Methode, die normalerweise für statistische Analysen eingesetzt wird, kennenlernen möchten. Ein Korpuslinguist wiederum wird die saubere mathematische Herleitung zu schätzen wissen. Wer sich noch allgemeiner mit Linguistik beschäftigt, erhält unter anderem eine gute Einführung in die Methoden der Korpuslinguistik. Ein weiterer Ansatzpunkt für künftige Forschungen könnte womöglich in der Diskussion um Konstruktionen im Rahmen der Konstruktionsgrammatik angelegt werden (cf. z. B. Fischer/Stefanowitsch 2007).

Literatur

- Busse, Dietrich und Wolfgang Teubert (1994): "Ist Diskurs ein sprachwissenschaftliches Objekt? Zur Methodenfrage der historischen Semantik". In: Busse, Dietrich/Hermanns, Fritz/Teubert, Wolfgang (Hrsg.): *Begriffsgeschichte und Diskursgeschichte. Methodenfragen und Forschungsergebnisse der historischen Semantik*. Opladen: 10–28.
- Fischer, Kerstin/Stefanowitsch, Anatol (eds.) (2007): *Konstruktionsgrammatik. Von der Anwendung zur Theorie*. Tübingen. (= *Stauffenburg Linguistik* 40).
- Foucault, Michel (1981): *Archäologie des Wissens*. Frankfurt a. M.
- Foucault, Michel (2000): *Die Ordnung des Diskurses*. Frankfurt a. M.

- Hermann, Michael/Leuthold Heiri (2008): *Politisches Spinnenprofil*. Online: http://www.sotomo.ch/index.php?option=com_content&view=article&id=57:sotomospider&catid=36:projekte&Itemid=57, Stand: xx.
- Mayr, Georg von (1877): *Die Gesetzmäßigkeit im Gesellschaftsleben. Statistische Studien*. München.
- Warnke, Ingo (2002): "Adieu Text – bienvenue Diskurs? Über Sinn und Zweck einer post-strukturalistischen Entgrenzung des Textbegriffs". In: Fix, Ulla et al. (eds.): *Brauchen wir einen neuen Textbegriff? Antworten auf eine Preisfrage*. Frankfurt a.M.: 125–141. (= *Forum Angewandte Linguistik* 40).