

Stefan Engelberg/Henning Lobin/Kathrin Steyer/Sascha Wolfer
**Muster, Dynamik, Komplexität –
eine Einführung in den Gegenstand
des Bandes**

1 Einleitung

In der Geschichte der Sprachwissenschaft hat das Lexikon in unterschiedlichem Maße Aufmerksamkeit erfahren. In jüngerer Zeit ist es vor allem durch die Verfügbarkeit sprachlicher Massendaten und die Entwicklung von Methoden zu ihrer Analyse wieder stärker ins Zentrum des Interesses gerückt. Dies hat aber nicht nur unseren Blick für lexikalische Phänomene geschärft, sondern hat gegenwärtig auch einen profunden Einfluss auf die Entstehung neuer Sprachtheorien, beginnend bei Fragen nach der Natur lexikalischen Wissens bis hin zur Auflösung der Lexikon-Grammatik-Dichotomie. Das Institut für Deutsche Sprache hat diese Entwicklungen zum Anlass genommen, sein aktuelles Jahrbuch in Anknüpfung an die Jahrestagung 2017 – „Wortschätze: Dynamik, Muster, Komplexität“ – der Theorie des Lexikons und den Methoden seiner Erforschung zu widmen.

2 Phänomene und Theorien

Eine Sprachbeschreibung – so lehrt uns der Blick in jede gut sortierte Buchhandlung – besteht üblicherweise aus einem „Grammatik“ genannten Handbuch und einem „Wörterbuch“ genannten lexikalischen Inventar. Die Grammatik beschreibt vor allem die syntaktischen und morphologischen Regularitäten der Sprache, der Schwerpunkt des Wörterbuchs liegt auf den Bedeutungen von Wörtern. Diese Zweiteilung wird auch in der Theoriebildung der Sprachwissenschaft in der Unterscheidung von Grammatik und Lexikon vollzogen, wobei es einen wesentlichen Unterschied zwischen den beiden Komponenten gibt. Zur Grammatik hat die Sprachwissenschaft eine große Zahl von Theorien hervorgebracht, zum Lexikon nicht.

Wirft man einen Blick zurück in die 1950er und 1960er Jahre, so findet man eine Fülle von Grammatiktheorien (bzw. zum Teil im engeren Sinne eher Syntaxtheorien), zum Beispiel die Phrasenstrukturgrammatik, die Transformationsgrammatik, die Dependenzgrammatik, die Kategorialgrammatik, die Stratifikati-

onsgrammatik, die Systemisch-funktionale Grammatik etc. (vgl. zu lexikalischen Aspekten verschiedener Grammatiktheorien **Müller**). Eine ähnliche Auflistung für Lexikontheorien lässt sich nicht erstellen. Die starke Fokussierung auf Grammatik bzw. Syntax ist wissenschaftshistorisch wohl nachzuvollziehen. Dennoch muss es etwas verwundern, wie wenig Aufmerksamkeit das Lexikon in vielen Perioden der modernen Sprachwissenschaft gefunden hat. Dabei hat das Lexikon viel zu leisten: Es konstituiert ein großes, aber kognitiv beherrschbares System von formgebundenen Bedeutungen, über die der Konnex zwischen Sprache und Welt hergestellt wird. Das System beinhaltet eine immens große Anzahl an internen formalen und semantischen Beziehungen zwischen lexikalischen Einheiten. Es verfügt über Mechanismen, die ein komplexes Wechselspiel zwischen lexikalischen Idiosynkrasien, Musterhaftigkeiten und Regeln garantiert, und es ist in hohem Maße durch Verfahren semantischer Anpassung und Flexibilität im Bestand gekennzeichnet.

Das alles war nur sehr eingeschränkt im Blickfeld der Linguistik der 1950er und 1960er Jahre. Einzelne lexikalische Aspekte wurden bearbeitet, etwa die Schnittstelle zwischen Wortbedeutungen und Syntax in der Theorie thematischer Rollen und in der generativen Semantik, das Verhältnis von Grammatik und lexikalischen Idiosynkrasien in der Valenztheorie oder semantische Strukturierungen im Wortschatz in der Prototypentheorie und der strukturalistischen Merkmals- und Wortfeldtheorie. Von einer umfassenden Lexikontheorie war die Linguistik aber noch weit entfernt. Am ehesten deuteten sich Ansätze dazu in der Valenztheorie mit ihrer zunehmenden Berücksichtigung semantischer Aspekte an.

Die Beobachtung, dass die empirische Adäquatheit formaler Grammatik- und insbesondere Syntaxtheorien nur allzu oft an den vermeintlichen Idiosynkrasien der in die Strukturen einzusetzenden Wörter scheiterte, führte dann ab den 1970er Jahren zu einer zunehmenden Beschäftigung mit dem Lexikon und den Eigenschaften von Wörtern. Eine wichtige Rolle spielte dabei die Hinwendung der formalen Semantik zum Lexikon. Vor allem Arbeiten zu Aspekt und Aktionsarten, die in den 1980er/90er Jahren dann unter anderem in ereignisstrukturelle Theorien der lexikalischen Semantik mündeten, wiesen hier den Weg zu einer expliziteren Darstellung lexikalischer Bedeutung. Davon profitierten auch dekompositionelle Ansätze, die unter anderem die Schnittstelle zwischen Grammatik und Lexikon modellierten, etwa die Konzeptuelle Semantik oder die Zwei-Ebenen-Semantik. In den 1970er Jahren formulierte Richard Carter (1978 [1988]), was Gegenstand einer Theorie des Lexikons sein müsse: lexikalische Inferenzen, semantische Relationen, das Wortkonzept, Lexikalisierungsprinzipien, das Argumentlinking, der Ablauf des Lexikonerwerbs und universelle Lexikalisierungstendenzen. Diese Ideen wurden in der Folgezeit unter anderem vom MIT-Lexikonprojekt aufgegriffen. Das Bedürfnis nach einer „Theorie des Lexikons“ drückte

sich dann Anfang der 1990er Jahre in der Gründung des gleichnamigen Sonderforschungsbereichs (Düsseldorf, Köln, Wuppertal) aus. Themen wie die Schnittstelle zur Grammatik und das Argumentlinking, das mentale Lexikon und lexikalische Verarbeitungsprozesse, aber auch Fragen der Repräsentation und der Regelmäßigkeit im Lexikon standen hier im Mittelpunkt (vgl. etwa Wunderlich (Hg.) 2006).

Mit der empirischen Wende, die im Übergang zum 21. Jahrhundert durch die Verfügbarkeit immer größerer und varietätenreicherer Korpora eingeleitet wurde, wird für die Lexikonforschung ein neues Kapitel aufgeschlagen. Dabei soll der Ausdruck „empirische Wende“ nicht vergessen machen, dass auch in den Dekaden zuvor vor allem im Bereich der Psycholinguistik viel empirisch anspruchsvolle und erkenntnisstiftende Arbeit insbesondere zu lexikalischen Verarbeitungsprozessen geleistet wurde. Dennoch dürfte es einleuchten, dass ein Bereich wie das Lexikon, das theoretische Linguisten über Jahrzehnte schon durch seinen schiereren Umfang erschreckt hat, durch einen empirischen Zugang über sprachliche Massendaten neue Aufmerksamkeit erfährt und Anlass gibt zu neuen Lexikonkonzepten (vgl. etwa **Schmid**).

Der korpuslinguistische Zugang zum Lexikon hat zum einen die Lexikografie mit einer neuen empirischen Basis versehen, wobei neben neuartigen Anwenderwörterbüchern auch große lexikalische Forschungsressourcen wie die WordNets und FrameNets entstanden sind. Zum anderen wird die empirische Entwicklung aber auch von einer sprachtheoretischen Debatte begleitet, in der die Abgrenzung zwischen Lexikon und Grammatik in Frage gestellt wird. Diese Debatte ist wesentlich über die Konstruktionsgrammatik angestoßen worden. Auch wenn man deren syntaktische Konzeption und ihre Skepsis gegenüber Bedeutungskomposition kritisch diskutieren mag, so muss man doch konstatieren, dass hier eine Sprachtheorie entwickelt wird, die vielen Anforderungen lexikontheoretischer Forschung, insbesondere bei der Musterbildung und dort, wo lexikalische Einheiten betroffen sind, die über die Wortgrenze hinausgehen, weit eher entgegenkommt als die Theorien früherer Jahrzehnte.

Mit der Verknüpfung konstruktionsgrammatischer Ideen und korpuslinguistischer Empirie treten in der lexikontheoretischen Forschung neue Konzepte in den Fokus. Gebrauchsfrequenzen, Kookkurrenzen und Assoziationswerte sind zu zentralen lexikalischen Daten geworden, und Phänomene werden in Form von Mustern und Konstruktionen modelliert. Einige Überlegungen zu den zentralen Konzepten des Musters, der Komplexität und der Dynamik wollen wir in Abschnitt 4 etwas vertiefen, bevor wir in Abschnitt 5 auf die verschiedenen Erkenntniswege eingehen, die die Beiträge verfolgen. Wir schließen diese Einleitung mit einigen Implikationen ab, die die Beiträge für die Sprachdokumentation haben. Zunächst wollen wir im nächsten Abschnitt jedoch den Aufbau des Bandes erläutern.

3 Aufbau des Bandes

Im vorliegenden Band werden die vorangegangenen Überlegungen in fünf Teilen vertieft. Der erste Teil thematisiert das Verhältnis von Lexikon und Grammatik: **Stefan Müller** (Berlin) stellt für regelbasierte Grammatikmodelle verschiedene Herangehensweisen zur Repräsentation und Integration lexikalischer Information dar, während **Hans C. Boas** (Austin) das mit dem Regelbegriff konkurrierende Konzept der Konstruktion auf seine Positionierung zwischen Lexikon und Grammatik hin überprüft. Diese beiden Beiträge entfalten das theoretische Spannungsfeld, in dem sich die Lexikonforschung heute befindet. **Barbara Stiebels** (Leipzig) zeigt an einem ausgewählten Phänomen darin exemplarisch, von welchen Parametern und in welcher subtiler Weise das Zusammenspiel lexikalischer Deutungen mit syntaktischen Strukturausprägungen gesteuert wird.

Der zweite Teil des Bandes, „Kookkurrenz und Konstruktion“, basiert auf dem Konzept der Kookkurrenz in Korpora und fragt danach, wie aus derartigen statistisch auffälligen Wortverwendungen wiederkehrende Konstruktionen abgeleitet werden können. **Stefan Th. Gries** (Santa Barbara) beantwortet diese Frage durch exemplarische Untersuchungen sowohl auf der Ebene phonologischer Merkmale von idiomatischen Ausdrücken als auch in Hinsicht auf das Auffinden von Mehrwort-Lexemen in Korpora generell. **Martin Hilpert** (Neuchâtel) geht bei der Erhellung dieses Zusammenhangs den umgekehrten Weg, indem er einen bestimmten Typ von Konstruktionen auf der Wortebene dahingehend analysiert, wie sich durch weitergehende statistische Analysen unterschiedliche Anwendungsbereiche für diese Konstruktion voneinander abgrenzen lassen. **Kathrin Steyer** und **Katrin Hein** (Mannheim) wenden sich der Produktivität von Konstruktionsmustern zu, die, ausgehend von usuellen Wortverbindungen, ein zuweilen erstaunliches Spektrum abgewandelter Verwendungsweisen im Bereich der Wortbildung aufweisen.

Den dritten Teil, „Kognition und Semantik“, leitet **Petra Schulz** (Frankfurt am Main) mit einem Beitrag ein, in dem der Erwerb von Verben bei Kindern als ein Prozess beschrieben wird, dem eine bestimmte, auf den Endzustand ausgerichtete Ereignisstruktur zugrunde liegt. Wie Petra Schulz betrachtet auch **Sabine Schulte im Walde** (Stuttgart) Partikelverben und zeigt für diese besonders produktive, bezüglich der Bedeutung des Partikels zugleich aber auch extrem ambige Wortklasse, wie Bedeutung und Argumentstruktur aus einem komplexen Zusammenspiel der einzelnen Bedeutungskomponenten hervorgehen. **Silvia Hansen-Schirra**, **Katharina Oster**, **Jean Nitzke** und **Anne-Kathrin Gros** (Mainz) befassen sich mit dem kognitiven Lexikon im Übersetzungsprozess und fokussieren dabei die Rolle von Kognaten. An ihnen zeigen sie, wie diese spezifischen kognitiven Verknüpfungen von zwei Sprachen auf Wortebene dazu beitragen, eine Übersetzung auf die Verbesserung des Verständnisprozesses hin zu optimieren.

Sebastian Löbner (Düsseldorf) schließlich unterzieht das Gebiet der formalen lexikalischen Semantik einer Revision, indem er zeigt, wie mit sogenannten Barsalou-Frames typische Derivations- und Kompositionsmuster mit diesem aus der Kognitionswissenschaft stammenden Konzept systematisch nachvollzogen werden können.

Im vierten Teil, „Komplexität und Dynamik“, werden zwei für die Tagung zentrale Begriffe aus vier unterschiedlichen Perspektiven aufgeschlüsselt. **Hans-Jörg Schmid** (München) schlägt ein soziokognitives Modell eines dynamischen Lexikons vor, das viele der in den drei vorangegangenen Teilen diskutierten Phänomene und ihre Deutungen zu integrieren erlaubt. Auch **Dirk Geeraerts** (Leuven) entwickelt ein Modell, das differenziert nach den Dimensionen der Standardisierung, der Formalisierung und der Homogenisierung sowohl diachrone Dynamik als auch synchrone Variation zu fassen erlaubt. In dem methodologisch ausgerichteten abschließenden Beitrag zeigen **Carolin Müller-Spitzer**, **Sascha Wolfer** und **Alexander Koplenig** (Mannheim) in zwei unterschiedlichen Studien, wie sich die Untersuchung von Sprachwandel trotz der Verfügbarkeit großer Korpora mit spezifischen statistischen Problemen konfrontiert sieht, wie aber auch mit vergleichsweise einfachen Mitteln grundlegende theoretische Annahmen der Linguistik zur gegenseitigen Abhängigkeit von Komplexität und Dynamik korpusanalytisch bestätigt werden können.

Im abschließenden fünften Teil des Bandes, „Wortschatz und Lexikografie“, wenden sich **Alexander Mehler**, **Rüdiger Gleim**, **Wahed Hemati** und **Tolga Uslu** (Frankfurt am Main) zunächst dem Zusammenhang zwischen lexikalischen und sozialen Netzwerken zu. Sie zeigen anhand des Wörterbuchs Wiktionary, wie sich die Dynamik seiner kollaborativen Bearbeitung in der Verteilung der lexikalischen Einheiten im Wörterbuch niederschlägt. Derartige soziale Dynamiken in Wiktionary bilden für **Christian M. Meyer** (Darmstadt) den Ausgangspunkt für die Untersuchung der Frage, wo die Stärken und Schwächen kollaborativer Wörterbücher liegen und wie die Lexikografie insgesamt von diesen Erfahrungen profitieren kann. Inwieweit die klassische Verlagslexikografie sich auf derartige und andere Herausforderungen der Digitalisierung einstellen kann, legt **Kathrin Kunkel-Razum** (Berlin) anhand der lexikografischen Arbeit im Dudenverlag dar. Nicht mehr andere Verlage sind die Hauptkonkurrenten in diesem sich stark verändernden Markt, sondern Angebote wie Wiktionary und Korrekturprogramme.

4 Dynamik, Muster und Komplexität

4.1 Dynamik

Sprachliche Dynamik wird in diesem Band in den unterschiedlichsten Formen konzeptualisiert und analysiert. Im Sinne einer Wachstumsdynamik greifen **Müller-Spitzer/Wolfer/Koplenig** Veränderungen im Wortschatz der deutschen Sprache auf: Sie zeigen am Beispiel der Verbbildungen mit gegen* und fremd*, wie schwierig es unter Umständen sein kann, „echte“ Wachstumsdynamik – also eine vermehrte Verwendung sprachlicher Elemente über die Vergrößerung und Ausweitung der Korpusgrundlage hinaus – von „artifiziellem“ Wachstum abzugrenzen. Letzteres tritt dann auf, wenn sich in einem Zeitraum die Größe und Zusammensetzung von Korpora so stark ändert (meist vergrößert), dass nicht mehr zweifelsfrei auf eine tatsächliche Ausbreitung sprachlicher Muster geschlossen werden kann. Unter anderem wird sprachliche Dynamik in diesem Beitrag auch als Produktivität verstanden. Dieser Auffassung schließt sich **Schulte im Walde** an. Sie zeigt etwa, wie Dynamik durch (teilweise neologistische) Kompositionsprozesse von Partikeln und Basisverben hervorgebracht wird.

Auch andere musterbezogene Produktivitätsprozesse können so verstanden werden. So zeigt **Hilpert** in seinem Beitrag, wie sprachliche Dynamik dadurch entsteht, dass verschiedene Teile eines Netzwerks aus sprachlichen Konstruktionen unterschiedlich produktiv sein können und zu spezifischen Verdichtungen im semantischen Raum einer sprachlichen Konstruktion führen (in seinem Falle bei englischen Partizipialkomposita). Ein solcher Vorgang kann als konkrete Instantiierung dessen verstanden werden, wenn **Schmid** Dynamik in seinem Beitrag als „inhärenten Wesenszug des Lexikons“ beschreibt. Das gilt nicht nur für mentale Lexika individueller Sprecher, sondern für das kollektive Lexikon einer ganzen Sprachgemeinschaft. Er widmet sich im Fortgang seines Beitrags dem Zusammenspiel stabilisierender und dynamisierender Kräfte, deren Analyse sprachliche Dynamik vorhersagbar machen soll.

Doch sprachliche Produktivität ist nicht das einzige Konzept, das mit Dynamik im Lexikon verbunden werden kann. Auf lexikografischer Seite (also dem Lexikon als sprachbeschreibendem und -dokumentierendem Instrument) zeigen sowohl **Mehler/Gleim/Hemati/Uslu** als auch **Meyer** in ihren Beiträgen, welche Rolle die inter-individuelle Dynamik bei der Erzeugung kollaborativ erarbeiteter lexikalischer Netzwerke spielt. Mehler et al. widmen sich hier eher der Kollaborationsdynamik, die zwischen den Mitwirkenden solcher lexikografischen Projekte herrscht, und unterziehen diese einer komputationalen netzwerkanalytischen Betrachtung. Meyer geht eher auf die dynamische Veränderung von Einträgen und

Instruktionen kollaborativer Wörterbücher ein und zeigt somit, welche Form von Dynamik entsteht, wenn in kollaborativen Wörterbüchern potenziell jede und jeder Beitragende jederzeit einen Eintrag verändern kann.

Geeraerts siedelt seine Betrachtungen wiederum auf einer anderen Ebene an, indem er darlegt, welche dynamischen Beziehungen zwischen verschiedenen Sprachebenen bestehen. Dabei bezieht er sich auf das Niederländische in den Niederlanden und Belgien und zeigt, welche zeitliche Dynamik zwischen den Sprachebenen durch Standardisierungsprozesse entsteht.

Schließlich erwächst aus dem Lexikon heraus auch eine interpretative Dynamik, die darin besteht, dass lexikalische Bedeutungen vielfältigen, oft metaphorischen oder metonymischen Reinterpretationsprozessen ausgesetzt sind, deren Ergebnisse selbst wieder lexikalisiert werden können. **Stiebels** etwa zeigt dies am Beispiel der Umdeutung (*coercion*) satzeinbettender Prädikate.

4.2 Muster

Seit einigen Jahren ist in Teilen der Syntaxforschung eine Abkehr von Phrasenstrukturmodellen zu verzeichnen, die ihren Ausgang genommen hat in strukturellen Untersuchungen realer Sprachdaten, dokumentiert in großen digitalen Korpora. Aufgrund der Varianz sprachlicher Strukturen, die in allen Dimensionen der Sprachverwendung ermittelt werden kann, geriet dabei der traditionelle Regelbegriff in die Kritik, da dieser zur Erfassung dieser Varianzphänomene nicht geeignet erschien. Parallel dazu hat sich auch die bereits seit längerer Zeit etablierte quantitative Auswertung digitaler Korpora weiter entfaltet. In dieser Forschungsströmung steht die Analyse von Frequenzen im Vordergrund: Wortfrequenzen, Bigramme, Kookkurrenzen. Korpuslinguistische Analysen dieser Art haben den Blick gelenkt auf die spezifischen kombinatorischen Eigenschaften, die einzelne Wörter in der Sprachverwendung besitzen und die bislang von der Grammatikforschung wie von der Lexikografie nur als ein Randphänomen betrachtet worden sind.

Beide Forschungsströmungen laufen in jüngerer Zeit im Konzept des Musters zusammen (vgl. Bücken 2015). Aus grammatischer Perspektive kennzeichnet ein Muster die Eigenschaft, mehrere Elemente miteinander in Verbindung zu setzen, ohne dabei Sequenz- oder Adjazenz-Bedingungen zu fixieren. Den wissenschaftshistorischen Prototyp eines syntaktischen Musters stellt die Verbvalenz dar: Durch das Konzept der Valenz werden bereits seit den 1960er Jahren die spezifischen Kombinationseigenschaften von Verben beschrieben und dabei neben der Systematisierung in Verbgruppen auch die kombinatorischen Besonderheiten bestimmter Verben untersucht und dokumentiert (vgl. Eroms 2003). Ohne eine kor-

puslinguistische Methodik gewann aber auch in der Valenzforschung die Tendenz zur Abstraktion der vorgefundenen Muster Oberhand, durch die schließlich in Gestalt von Subkategorisierungsrahmen Muster auch in Phrasenstrukturmodelle Einzug erhalten haben.

Aus lexikalischer Sicht sind Muster neben einer mehr oder weniger starken Abstraktion von Sequenzeigenschaften oft auch durch eine Alternation von festen lexikalischen Elementen und mehr oder weniger stark restringierten abstrakten Einheiten (Slots) gekennzeichnet, wie etwa Wortbindungsmuster, Argumentstrukturmuster oder Wortbildungsmuster (siehe etwa **Steyer/Hein, Hilpert, Löbner**). Dabei drückt sich in den Frequenzen und der Varianz der Slotbesetzungen die unterschiedliche Produktivität solcher Muster aus.

Aus korpuslinguistischer Perspektive geht es in der Operationalisierung des Konzepts des Musters weniger um die zunehmende Erhöhung des Abstraktionsgrades als um die Verfeinerung der statistischen Verfahren ihrer Erkennung. Insbesondere die Kookkurrenzanalyse hat es möglich gemacht, in großen Korpora sehr differenzierte, spezifische Muster zu ermitteln, die die im Text vorgefundenen Wort-Types umgeben (vgl. Perkuhn/Keibel/Kupietz 2012, S. 116 ff.). Derartige Muster lassen sich zwar nicht immer linguistisch deuten, denn sie spiegeln auch semantische Eigenschaften der Textdomäne wider, doch erlauben es die Verfahren der Kookkurrenzanalyse zu bestimmen, wieviel grammatische und semantische Struktur Wörter jeweils mit sich führen und wie sich das Zusammenspiel dieser Wortstrukturen in größeren Mustern verfestigen kann.

Der Begriff des Musters ist also deutlich durch die Perspektive auf den Sprachgebrauch geprägt und methodisch mit der empirischen Untersuchung von Korpora verbunden. Dabei treten theoretische Betrachtungen in den Hintergrund. Indifferent steht dieser Begriff insbesondere der Art der Konstituierung einer übergreifenden „Musterbedeutung“ gegenüber – Mustern kann sowohl kompositionell als auch nicht-kompositionell – idiosynkratisch – eine Bedeutung zugeschrieben werden. Die Gruppe der idiosynkratisch konstituierten Muster ist in Gestalt der Konstruktion Gegenstand eines alternativen grammatiktheoretischen Paradigmas geworden (vgl. als Übersicht Ziem/Lasch 2013 und hier im Band **Boas** versus **Müller**). Die Konstruktionsgrammatik kann als die Verbindung einer gebrauchsbasierten, den klassischen Regelbegriff ablehnenden Syntaxauffassung mit der korpuslinguistischen Analyse des Wortgebrauchs verstanden werden. Diese Verbindung wird auch terminologisch zum Ausdruck gebracht, wenn von dem Bestand der Konstruktionen einer Sprache als „Konstruktikon“ gesprochen wird (Croft/Cruise 2004) oder von den aus Kookkurrenzanalysen ableitbaren Konstruktionen als „Kollostruktionen“ (Stefanowitsch/Gries 2003). Mit der Verabsolutierung des Konstruktionsbegriffs werden alle Arten von Mustern erfasst, also auch solche, in denen eher eine kompositionelle Bedeutungskonstitution zu ver-

zeichnen ist. Kompositionalität wird dabei nicht als der Normalfall der Bedeutungskonstitution verstanden, sondern als ein Spezialfall, der wiederum durch übergeordnete abstrakte Konstruktionen gesteuert wird.

4.3 Komplexität

In jüngerer Zeit tauchte mit dem Revival einer scheinbar ad acta gelegten Frage, nämlich, ob alle Sprachen eigentlich gleich komplex seien, das Bedürfnis auf, den Komplexitätsbegriff zu operationalisieren (vgl. etwa Sampson 2009). Dazu ist mittlerweile eine reichhaltige Literatur entstanden, die allerdings vor allem auf grammatische Komplexität Bezug nimmt. Das Lexikon ist kaum zum Gegenstand dieser Diskussion geworden.

Diskussionen über sprachliche Komplexität setzen an verschiedenen Punkten an: Wie leicht ist ein sprachliches System zu beschreiben, zu erzeugen, zu lernen, welchen Grad an Organisiertheit und dynamischem Potenzial weist es auf? Die Ansätze zur Operationalisierung von Komplexität variieren dabei stark (vgl. Szmrecsanyi/Kortmann 2012). Maße für grammatische Komplexität können auf der einen Seite phänomenorientiert auf dem Auszählen von Kategorien und ihren Ausprägungen, paradigmatischen Klassen und ihren Varianten, Arten von syntagmatischen Abhängigkeiten etc. basieren (z.B. Nichols 2009) oder auf der anderen Seite sehr abstrakt auf informatischen Maßen, die etwa die strukturelle Komplexität einer Zeichenkette über das kürzeste Programm definiert, das zu ihrer Beschreibung erforderlich ist (Kolmogorow-Komplexität).

Betrachtet man sprachliche Komplexität nicht als Ganzes, sondern differenziert zwischen der Komplexität von Subsystemen, so sind oft Trade-Off-Effekte zu beobachten. Bekannt ist die Annahme, dass eine Sprache, die wortstrukturell komplex ist, in der Wortstellung wenig restriktiv ist. Korpusbasiert versucht man, solche Trade-Offs über informationstheoretische Komplexitätsmaße nachzuvollziehen (vgl. **Müller-Spitzer/Wolfer/Koplenig**).

Auch wenn man lexikalische Komplexität vermutlich nicht über die bloße Menge lexikalischer Einheiten erfassen möchte, so sei doch darauf hingewiesen, dass die größten Korpora des Deutschen mittlerweile Einwortlexeme wohl im zweistelligen Millionenbereich belegen (Engelberg 2015); dabei sind Mehrwortlexeme und Lesartendisambiguierung noch nicht berücksichtigt. Das ist jedenfalls der Bestand an Einheiten, mit denen es Strukturen, Verarbeitung und Expansionsprozesse im Lexikon zu tun haben.

Überlegungen zu lexikalischer Komplexität können an verschiedenen Stellen ansetzen. Relationale Komplexität – als Komplexität des zugrundeliegenden Graphen – kann die Menge an Relationen ins Visier nehmen, die zwischen lexika-

lischen Einheiten bestehen, etwa semantische Relationen wie Antonymie, Hyperonymie, Meronymie, morphologische Relationen zwischen Komponenten komplexer Wörter, aber auch die Anzahl an Lesarten, die Lexeme an sich binden. Die Menge der etablierten Relationen kann dabei als Komplexitätsindikator genommen werden, besser noch die durchschnittliche Menge an Relationen, über die ein Lexem verfügt, und die einen Eindruck von der Dichte des lexikalischen Netzes vermittelt (Müller-Spitzer/Wolfer 2015; siehe auch **Mehler/Gleim/Hemati/Uslu**). Dabei sind auch hier Trade-Off-Effekte zu beobachten. Die morphologische Komplexität von Wörtern korreliert negativ mit ihrer Gebrauchshäufigkeit, und der Grad an Polysemie von Lexemen korreliert positiv mit der Häufigkeit des Lexems. Entsprechend korreliert der Grad an Polysemie negativ mit der morphologischen Komplexität von Wörtern: Je polysemer Wörter sind, um so morphologisch einfacher sind sie; semantisch komplexe Wörter tendieren zu geringer morphologischer Komplexität und umgekehrt.

Das Lexikon ist ein inhärent dynamisches System (vgl. **Schmid**). Will man Komplexität an Dynamik und Wandel knüpfen, wie Larsen-Freeman (2012) es vorschlägt, so kann man den Komplexitätsbegriff an die produktiven Expansionsprozesse des Lexikons knüpfen. Dazu gehören etwa konkatenative morphologische Prozesse (siehe etwa **Schulte im Walde**) und koerzive semantische Anpassungen (siehe etwa **Stiebels**). Immer deutlicher wird aber auch, wie sehr lexikalische Kreativität durch eher analogisch wirkende Prozesse in einem Netzwerk von Mustern bestimmt ist. Diese wirken nicht nur im Wortbildungsbereich (vgl. **Hilpert**), sondern auch hinsichtlich der Assoziation von Lexemen mit syntaktischen Konstruktionen (etwa im Argumentstrukturbereich) oder bei der Ausbeutung von Mustern im Bereich usueller Wortverbindungen (**Steyer/Hein**). Wenn wir von „kreativer“ Komplexität des Lexikons sprechen wollen, so orientiert sich diese an der Anzahl und Produktivität lexikalischer Muster.

5 Erkenntniswege

Ebenso vielfältig wie die in den Beiträgen behandelten Themen sind die empirisch-methodischen Zugangsweisen zu den jeweiligen linguistischen Phänomenen. Daten aus einer weiten Bandbreite linguistischer Korpora nehmen in einigen Beiträgen nach wie vor eine zentrale Rolle ein. **Hilpert** geht dem Phänomen der Produktivität anhand eines englischsprachigen diachronen Korpus nach. **Müller-Spitzer/Wolfer/Koplenig** nähern sich einerseits Phänomenen des Wortschatzwandels, wie sie in den Korpora des Instituts für Deutsche Sprache (IDS) belegt sind, andererseits zeigen sie anhand eines Korpus von Übersetzungen der Bibel,

wie sprachvergleichende Untersuchungen möglich sind. **Steyer/Hein** bedienen sich ebenfalls der Korpora des IDS und beschäftigen sich mit mehr oder weniger festen Wortverbindungen. **Stiebels** nutzt für ihre Studie zur Polysemie bei satzeinbettenden Prädikaten eine Kombination verschiedener Korpora sowie eine eigens auf ihre Fragestellung hin optimierte Datenbank.

Korpora werden aber auch immer häufiger durch Informationen aus anderen Quellen ergänzt. So kombinieren sowohl **Gries** als auch **Schulte im Walde** korpuslinguistische Herangehensweisen mit Experimentaldaten von Sprecherinnen und Sprechern. Ein solcher Ansatz, der auch unter dem Stichwort „konvergierende Evidenz“ firmiert, macht es möglich, sich einem linguistischen Phänomen aus mehreren methodischen Richtungen zu nähern. Eine solche „Datentriangulation“ gewinnt in der linguistischen Forschung zusehends an Einfluss, da es wissenschaftlich äußerst attraktiv ist, ein Phänomen aus unterschiedlichen empirischen Blickwinkeln zu beleuchten, um so Konvergenzen oder Divergenzen festzustellen. Einen ähnlichen Ansatz verfolgen auch **Hansen-Schirra/Oster/Nitzke/Gros**, indem sie Blickbewegungs- und Key-Logging-Daten kombinieren.

Auch die linguistische Erschließung eher ungewöhnlicher bzw. ungewohnter Datenquellen wird in einigen Beiträgen beschrieben. So nutzen sowohl **Mehler/Gleim/Hemati/Uslu** als auch **Meyer** freie lexikografische Ressourcen, allen voran das „Wiktionary“, um den Prozessen und Mechanismen auf den Grund zu gehen, die während der Erstellung dieser Ressourcen wirken. Auch **Geeraerts** verwendet für seine Studie eine selten erschlossene Quelle linguistischer Forschung, nämlich Preisschilder in lokalen Bekleidungsgeschäften, um seinen lexikalisch-pektometrischen Ansatz zur Unterscheidung zwischen Demotisierung und Destandardisierung in den Nationalvarietäten des Niederländischen zu unterfüttern.

Einige Beiträge widmen sich zudem explizit einer meta-methodischen Perspektive. So stellt **Gries** mehrere Maße vor, die häufiger in korpuslinguistischen Untersuchungen herangezogen werden und reflektiert deren Möglichkeiten und Grenzen. **Müller-Spitzer/Wolfer/Koplenig** zeigen u.a. auf, welche hohen Anforderungen an linguistische Datensammlungen bestehen, wenn Wortschatzwandelphänomene untersucht werden sollen. Auch ihr Beitrag kann als eine Auslotung der Grenzen und Möglichkeiten empirisch-linguistischer Fragestellungen verstanden werden.

Für einige Autoren steht nicht die konkrete Analyse eines bestimmten Datensatzes im Vordergrund. Dafür können die jeweiligen Beiträge aber als Ausgangspunkt für empirisch-quantitative Untersuchungen dienen. So skizziert **Schmid** ein kohärentes soziokognitives Modell des dynamischen Lexikons, das aus einer breiten Basis experimentell belegter Effekte heraus motiviert ist. Das Modell als Ganzes kann dabei wiederum zur Generierung neuer, empirisch überprüfbarer

Hypothesen genutzt werden. Auch der Beitrag von **Löbner**, der für die Verwendung von Barsalou-Frames in formalen Ansätzen der Semantik plädiert, kann unter diesem Aspekt gesehen werden. So sind dem dargestellten Modell Prinzipien (wie bspw. der Prozess der *coercion*) inhärent, die hypothesengenerierenden Charakter haben. **Müllers** Ausführungen zum Verhältnis zwischen Lexikon und Grammatik und deren Formalisierung in verschiedenen Grammatiktheorien eignen sich ebenso zur Ableitung von Vorhersagen, die in einem weiteren Schritt empirisch überprüft werden können.

6 Sprachdokumentation

Die wissenschaftliche Lexikografie sieht sich seit nunmehr fast zwei Jahrzehnten neuen Entwicklungen gegenüber, die zu teils einschneidenden Veränderungen geführt haben und zu Neupositionierungen zwingen (siehe **Kunkel-Razum; Meyer; Mehler/Gleim/Hemati/Uslu**) (vgl. Klosa/Müller-Spitzer (Hg.) 2016). Man kann dies in drei Punkten zusammenfassen: a) innovative Zugänge zu authentischen sprachlichen Massendaten – bedingt durch die sprachtechnologische Revolution – und die damit verbundenen neuen (Muster-)Sichten auf das Lexikon sowie auf alle Aspekte des Sprachgebrauchs (siehe **Boas; Gries; Hilpert; Steyer/Hein**); b) neue Formen der lexikografischen Repräsentation – eine Abkehr von der Dominanz narrativer (Meta-)Beschreibungen und eine Hinwendung zu intelligent systematisierten Daten, Graphen und Netzwerken (siehe **Möhrs/Müller-Spitzer** zu den Projekten der Methodenmesse), zu multimedialen und -modalen Datenbanken und sprachbezogenen Forschungsressourcen (siehe **Boas**); c) das Internet als enzyklopädisches Wissensuniversum, bei dem sich die Teilhabe an lexikografischen – und generell an sprachdokumentarischen Prozessen – immer mehr von den etablierten Redaktionsstuben der Wörterbuchverlage und Akademien in eine breite, partizipierende Community verlagert (siehe **Meyer; Mehler/Gleim/Hemati/Uslu**).

In Bezug auf die empirische Fundierung durch sprachliche Massendaten lässt sich festhalten, dass Korpora mittlerweile auch in den lexikografischen Werkstätten angekommen sind, zum einen in Form großer elektronischer Belegsammlungen, zum anderen bei der Anwendung automatischer Analyseverfahren wie Frequenz- und Kookkurrenzanalysen, die als wichtige heuristische Werkzeuge für die Überprüfung der Usualität der zu beschreibenden Einheiten und für einen kontextualistischen Zugang zu deren Bedeutung und Gebrauch dienen. Eine besondere Herausforderung stellen die neuesten linguistischen Erkenntnisse über sprachliche Verfestigung und Musterhaftigkeit als genuine Sprachprinzipien dar

(siehe **Schmid, Steyer/Hein**). Es ist beispielsweise zu fragen, wie lexikalisch partiell gefüllte Schemata – und vor allem ihre funktionalen Restriktionen – Sprachnutzern, z.B. im Fremdsprachenbereich, nahegebracht werden sollten. Die grundsätzliche Unterscheidung zwischen stark lexikalisierten Schemata, die als solche erworben werden müssen, einerseits und schematisierten Konstruktionen, bei denen die Art und Weise der Slotbesetzungen auf verschiedenen Abstraktions Ebenen zu verstehen ist, andererseits, führt zwangsläufig auch zu neuartigen (muster-)lexikografischen Beschreibungsformaten. Solche Formate können beispielsweise intelligente Inventarisierungen und Hierarchisierungen sprachlicher Einheiten und ihrer kontextuellen Einbettungsmuster sein; mithilfe von Statistiken, Kookkurrenz- und Füllertabellen, systematisierten Korpusbelegen und vernetzt in Online-Portalen und Wörterbuchverbänden. Die Chance, authentische Sprachdaten in großem Ausmaß ins Zentrum der Beschreibung zu rücken, ist gleichzeitig aber auch die größte Herausforderung: Denn es wird für Nutzer nicht leichter, sich in der Flut von Daten zu orientieren und das Erklärungspotenzial, aber auch die Grenzen des jeweils angebotenen sprachdokumentarischen Produkts zu erfassen. Die Notwendigkeit einer wie auch immer gearteten Didaktisierung ist – bei aller Euphorie angesichts des neuen Wissensuniversums – daher wieder stärker ins Bewusstsein zu rücken. Den Forderungen nach Methodentransparenz und einer Reflektion über die Nachvollziehbarkeit der dargestellten Ergebnisse kann und muss sich in erster Linie die wissenschaftliche Lexikografie stellen.

Korpusanalysen unterstützen aber nicht nur die Quantität der lexikografischen Inhalte, sondern ermöglichen auch die qualitative Erfassung feiner, distinktiver Gebrauchsnuancen, die eine kulturell angemessene Sprachkompetenz ganz wesentlich ausmachen. Derart tiefergehende Beschreibungen lassen sich aber realistischerweise nur für kleinere Sprachausschnitte durchführen. Nützlich im Sinne des Nachschlagenden sind jedoch vor allem aufbereitete Sprachdaten, die für den jeweiligen Phänomenbereich einen gewissen Grad an Vollständigkeit beanspruchen, eine Quadratur des Kreises – und das vor dem Hintergrund sich rasant entwickelnder Internetportale und Informationssysteme, mit denen die wissenschaftliche Lexikografie aus quantitativer Sicht in der Tat nicht Schritt halten kann. In kollaborativen Erarbeitungsszenarien können unter der Teilhabe vieler Nutzer in sehr viel kürzerer Zeit sehr viel umfassendere Daten erfasst werden (siehe **Mehler/Gleim/Hemati/Uslu**). Ein grundsätzliches Problem bleibt aber der fehlende Qualitätsstandard bezogen auf die empirische Validierung, z.B. die Einbeziehung von Korpusdaten oder anderer wissenschaftlich belegter Quellen (siehe **Meyer**). Vor diesem Hintergrund stellen sich für die wissenschaftliche Lexikografie folgende Fragen: Inwiefern sollte sie sich als Teil der kollaborativen Bewegung verstehen? Ist es möglich, dass sie sich für die Notwendigkeit

eines soliden methodisch-empirischen Fundaments in der Autoren- und Nutzer-gemeinde Gehör verschafft und zumindest eine Sensibilisierung erreichen kann? Und wenn ja, wie könnte eine Verbindung von „kollektiver Weisheit“ und etabliertem Fachwissen produktiv ausgestaltet werden?

Mit dem vorliegenden Jahrbuch des Instituts für Deutsche Sprache werden diese Fragen nicht endgültig zu beantworten sein. Wir hoffen aber, dass die in diesem Band zusammengestellten Beiträge das Gesamtbild einer gegenwärtig hochinteressanten Entwicklung im Bereich der Lexikologie erkennen lassen, das auch eine Vielzahl neuartiger Erklärungen bietet. Der Wortschatz einer Sprache mit den in ihm enthaltenen Mustern, seiner Dynamik und seiner Komplexität scheint in der linguistischen Forschung endlich den Stellenwert erlangt zu haben, der ihm in der alltäglichen Verwendung von Sprache schon immer zugekommen ist.

Literatur

- Bücker, Jörg (2015): Schema – Muster – Konstruktion. In: Dürscheid, Christa/Schneider, Jan Georg (Hg.): Handbuch Satz, Äußerung, Schema. (= Handbücher Sprachwissen 4). Berlin u.a., S. 445–463.
- Carter, Richard (1978 [1988]): Arguing for semantic representations. In: *Recherches Linguistiques 5–6* (Université de Paris VIII), S. 61–92. [Auch in: Levin, Beth/Tenny, Carol (Hg.): *On linking: Papers by Richard Carter*. Cambridge, MA, S. 139–166].
- Croft, William/Cruise, David A. (2004): *Cognitive linguistics*. Cambridge.
- Engelberg, Stefan (2015): Quantitative Verteilungen im Wortschatz. Zu lexikologischen und lexikografischen Aspekten eines dynamischen Lexikons. In: Eichinger, Ludwig M. (Hg.): *Sprachwissenschaft im Fokus. Positionsbestimmungen und Perspektiven*. (= Jahrbuch des Instituts für Deutsche Sprache 2014). Berlin/Boston, S. 205–230.
- Eroms, Hans-Werner (2003): Die Wegbereiter einer deutschen Valenzgrammatik. In: Ágel, Vilmos et al. (Hg.): *Dependenz und Valenz/Dependency and valency*. Halbbd. 1. (= Handbücher zur Sprach- und Kommunikationswissenschaft 25.1). Berlin/New York, S. 159–169.
- Klosa, Annette/Müller-Spitzer, Carolin (Hg.) (2016): *Internetlexikografie. Ein Kompendium*. Berlin/Boston.
- Larsen-Freeman, Diane (2012): Complex, dynamic systems. A new transdisciplinary theme for applied linguistics? In: *Language Teaching* 45, S. 202–214.
- Müller-Spitzer, Carolin/Wolfer, Sascha (2015): Vernetzungsstrukturen digitaler Wörterbücher. Neue Ansätze zur Analyse. In: *Lexicographica* 31, S. 173–199.
- Nichols, Johanna (2009): Linguistic complexity: A comprehensive definition and survey. In: Sampson/Gil/Trudgill (Hg.), S. 110–125.
- Perkuhn, Rainer/Keibel, Holger/Kupietz, Marc (2012): *Korpuslinguistik*. Paderborn.
- Sampson, Geoffrey (2009): A linguistic axiom challenged. In: Sampson/Gil/Trudgill (Hg.), S. 1–18.
- Sampson, Geoffrey/Gil, David/Trudgill, Peter (Hg.) (2009): *Language complexity as an evolving variable*. Oxford u.a.

- Szmrecsanyi, Benedikt/Kortmann, Bernd (2012): Introduction. Linguistic complexity. Second language acquisition, indigenization, contact. In: Kortmann, Bernd/Szmrecsanyi, Benedikt (Hg.): Linguistic complexity. Second language acquisition, indigenization, contact. (= Language & Litterae 13). Berlin/Boston, S. 6–34.
- Stefanowitsch, Anatol/Gries, Stefan Th. (2003): Collostructions: Investigating the interaction between words and constructions. In: International Journal of Corpus Linguistics 8, 2, S. 209–243.
- Wunderlich, Dieter (Hg.) (2006): Advances in the theory of the lexicon. (= Interface Explorations 13). Berlin/New York.
- Ziem, Alexander/Lasch, Alexander (2013): Konstruktionsgrammatik. Konzepte und Grundlagen gebrauchsbasierter Ansätze. (= Germanistische Arbeitshefte 44). Berlin/Boston.