



**ZDL**

Zentrum für digitale Lexikographie  
der deutschen Sprache

# Paradigmenwechsel in der Lexikographie

## Herausforderung und Chance

Vorträge zum Auftakt des Zentrums für  
digitale Lexikographie der deutschen Sprache (ZDL)

### **Herausgeber**

Die Präsidenten der Berlin-Brandenburgischen  
Akademie der Wissenschaften und der Akademie  
der Wissenschaften zu Göttingen



PARADIGMENWECHSEL IN DER LEXIKOGRAPHIE  
Herausforderung und Chance

Reden zur Auftaktveranstaltung des Zentrums für digitale  
Lexikographie der deutschen Sprache am 1. Januar 2019 an der  
Berlin-Brandenburgischen Akademie der Wissenschaften



# PARADIGMENWECHSEL IN DER LEXIKOGRAPHIE

## HERAUSFORDERUNG UND CHANCE

---

Die Präsidenten der Berlin-Brandenburgischen Akademie der  
Wissenschaften und der Akademie der Wissenschaften zu Göttingen  
Herausgeber

Herausgeber: Die Präsidenten der Berlin-Brandenburgischen Akademie der  
Wissenschaften und der Akademie der Wissenschaften zu Göttingen

Redaktion: Julia Naji, Bernhard Roll

Grafik: angenehme gestaltung/Thorsten Probst

Druck: Druckerei Conrad, Berlin

© Berlin-Brandenburgische Akademie der Wissenschaften, 2019

Jägerstraße 22–23, 10117 Berlin, [www.bbaw.de](http://www.bbaw.de)

Nachdruck, auch auszugsweise, nur mit ausdrücklicher Genehmigung der Herausgeber.

ISBN: 978-3-939818-87-8

# INHALTSVERZEICHNIS

<b>Vorwort</b> .....	6
<b>Begrüßung</b>	
Martin Grötschel, Präsident der BBAW .....	8
<b>Grußworte und Einführung</b>	
Andreas Gardt, Präsident der AdWG .....	13
Thomas Rachel, Staatssekretär BMBF .....	16
Steffen Krach, Staatssekretär für Wissenschaft und Forschung, Berlin .....	20
<b>Gegenwart und Zukunft der deutschen Lexikographie</b>	
Wolfgang Klein, Akademiemitglied der BBAW, Gesamtkoordinator des ZDL .....	23
<b>Den Wortgebrauch verlässlich und umfassend beschreiben: die Ziele der ZDL-Plattform</b>	
Alexander Geyken, Leiter der Berliner Arbeitsstelle des ZDL .....	35
<b>„Wortgeschichte digital“: Aufgaben und Umsetzungsstrategien</b>	
Volker Harm, Leiter der Göttinger Arbeitsstelle des ZDL .....	47
<b>Neologismen im Kontext der lexikalischen Forschung am Leibniz-Institut für Deutsche Sprache</b>	
Henning Lobin, Leibniz-Institut für Deutsche Sprache .....	55
<b>Variatio delectat? Zum Verhältnis von Lexikographie und sprachlicher Vielfalt</b>	
Alexandra Lenz, Österreichische Akademie der Wissenschaften, Universität Wien .....	64
<b>Sprachpolitik ist Kulturpolitik: Gegenwartslexikographie als öffentliche Dienstleistung in Skandinavien</b>	
Lars Trap-Jensen, Ordnet.dk .....	89
<b>Digital Lexicography and Oxford Dictionaries: The International Context</b>	
Sarah Ogilvie, OUP Dictionaries .....	96
<b>Abendvortrag mit anschließendem Gespräch Die deutsche Sprache, ihr Wortschatz und internationale Spracharbeit</b>	
Klaus-Dieter Lehmann, Goethe-Institut, Akademiemitglied BBAW .....	101

## VORWORT

Wörterbücher stehen in jedem Bücherregal und liegen auf vielen Schreibtischen. Wer bewusst schreibt und redet, wer sich für Sprache und deren Gebrauch interessiert, kommt ohne Wörterbücher nicht aus. Die Errungenschaften der letzten 1200 Jahre Wörterbuchgeschichte verdienen größte Anerkennung. Die Sorgfalt und Ausdauer, welche die lexikographische Arbeit in der Vergangenheit ausgezeichnet hat, ist ganz und gar bemerkenswert. Aber mit dem technischen Fortschritt verändern sich die Nutzeransprüche, die Wörterbucharbeit der Gegenwart und Zukunft muss diesen neuen Erwartungen entsprechen, ihr stehen aber auch ganz neue Möglichkeiten zur Verfügung. Die Lexikographie der Zukunft, es kann daran kein Zweifel mehr bestehen, ist digital. Das ist das Thema dieses Sammelbandes.

Die wissenschaftlichen Akademien sind ihrer Tradition nach Gelehrtenvereinigungen, deren Ziel es ist, Wissen zu schaffen und kulturelles Erbe zu bewahren. Aber wie jede wissenschaftliche Tätigkeit sollte auch diese nicht nur den Gelehrten selbst zugutekommen, sondern der gesamten Öffentlichkeit. Vier deutsche Akademien – die Akademie der Wissenschaften zu Göttingen, die Akademie der Wissenschaften und der Literatur in Mainz, die Sächsische Akademie der Wissenschaften zu Leipzig und die Berlin-Brandenburgische Akademie der Wissenschaften gemeinsam mit dem Leibniz-Institut für Deutsche Sprache als Kooperationspartner haben sich zusammengetan, um ein neues Großprojekt auf den Weg zu bringen, das dieser Verantwortung in Zukunft ganz unmittelbar nachzukommen versucht: Das Zentrum für digitale Lexikographie der deutschen Sprache (ZDL), ein digitales lexikalisches System, das den deutschen Wortschatz in Geschichte und Gegenwart wissenschaftlich fundiert und umfangreich darstellt und das seine Erkenntnisse für alle über das Internet kostenlos verfügbar macht. Gefördert wird das Vorhaben durch das Bundesministerium für Wissenschaft und Forschung, wesentliche Teile von bereits im Akademienprogramm geförderten Wörterbüchern werden ihre Ergebnisse einbringen.

Das ZDL hat am 1. Januar 2019 seine Arbeit aufgenommen. Am 29. Januar 2019 fand im Leibniz-Saal der Berlin-Brandenburgischen Akademie der Wissenschaften der feierliche Auftakt statt, zu dem rund 250 Gäste aus dem In- und Ausland,

darunter große Persönlichkeiten aus der Politik und der Wissenschaft, sowie zahlreiche Pressevertreter zusammengekommen sind. Über das sichtbar große Interesse seitens der wissenschaftlichen, aber auch der allgemeinen Öffentlichkeit haben wir uns besonders gefreut. Um diesem feierlichen Tag ein kleines Denkmal zu setzen, haben wir diesen Band zusammengestellt. Er enthält alle Vorträge, die zum Auftakt des ZDL präsentiert worden sind, ebenso wie einige fotografische Eindrücke, welche die Atmosphäre im Leibniz-Saal widerspiegeln. Wir danken allen, die zum Gelingen der Auftaktveranstaltung und zu diesem Sammelband beigetragen haben.

*Andreas Gardt*

*Martin Grötschel*

*Wolfgang Klein*

**Martin Grötschel** ist Präsident der Berlin-Brandenburgischen Akademie der Wissenschaften. Er war bis 2015 Mathematikprofessor an der Technischen Universität Berlin und Präsident des Konrad-Zuse-Zentrums für Informationstechnik Berlin. Seine Forschungsschwerpunkte liegen in der mathematischen Optimierung, der diskreten Mathematik und dem Operations Research. Schon in den 1990er Jahren interessierte er sich für die Verknüpfung von Informatik und Geisteswissenschaften. Daraus erwuchs sein besonderes Engagement für die Digital Humanities.



Foto: BBAW, Judith Affolter

Meine sehr verehrten Damen und Herren,

wir versammeln uns heute hier, um die Eröffnung des Zentrums für digitale Lexikographie der deutschen Sprache zwar nicht heftig zu feiern, aber doch wissenschaftlich angemessen zu begehen. Mit Worten sparsam umgehend, werde ich das Zentrum für digitale Lexikographie der deutschen Sprache ab jetzt einfach kurz ZDL nennen.

Ich möchte zu Beginn dieser Auftaktveranstaltung vier Personen herzlich begrüßen: den Vertreter des Bundesministeriums für Bildung und Forschung, Herrn Staatssekretär Thomas Rachel – das BMBF hat die finanzielle Förderung des ZDL für die nächsten fünf + drei Jahre übernommen und eine Verstetigung in Aussicht gestellt – den Berliner Staatssekretär für Wissenschaft und Forschung, Herrn Steffen Krach, den Präsidenten der Göttinger Akademie der Wissenschaften, Andreas Gardt – er wird im Anschluss das Wort ergreifen auch für die drei anderen Akademien, die gemeinsam mit der BBAW das ZDL betreiben werden – und das Mitglied der Berlin-Brandenburgischen Akademie der Wissenschaften und herausragenden Linguisten Wolfgang Klein, der die Koordination des ZDL übernommen hat.

Meine Damen und Herren, liebe Kolleginnen und Kollegen, als Hausherr und Präsident der Berlin-Brandenburgischen Akademie der Wissenschaften darf ich Sie alle sehr herzlich hier am Gendarmenmarkt im Leibniz-Saal unserer Akademie zur ZDL-Auftaktveranstaltung willkommen heißen. Bei der heutigen Veranstaltung handelt es sich um eine gemeinsame Veranstaltung unserer Akademie und der Akademie der Wissenschaften zu Göttingen (AdWG).

Zum 1. Januar dieses Jahres hat das Zentrum für digitale Lexikographie der deutschen Sprache seine Arbeit aufgenommen. Das Ziel des ZDL ist es, mit Hilfe eines digitalen Informationssystems den deutschen Wortschatz umfassend und verlässlich zu beschreiben. Dabei wird nicht nur der Wortschatz der Gegenwart wissenschaftlich erfasst, es werden auch die Sprachvarianten vergangener Jahrhunderte berücksichtigt. Besonders hervorheben möchte ich, dass diese digitale Plattform im Internet für jeden Interessierten frei und kostenlos zugänglich sein wird. Dieses herausragende Projekt steht unter der Schirmherrschaft der Union der deutschen Akademien der Wissenschaften, deren Präsidenten Hanns Hatt ich ebenfalls hier begrüßen darf. Das ZDL wird von der Berlin-Brandenburgischen Akademie der Wissenschaften, die die Koordination des Projektes übernommen hat, der Göttinger Akademie der Wissenschaften, der Mainzer Akademie der Wissenschaften und Literatur sowie der Sächsischen Akademie der Wissenschaften getragen. Und an dieser Stelle möchte ich auch den Präsidenten der Sächsischen Akademie der Wissenschaften Hans Wiesmeth willkommen heißen. Beim ZDL handelt es sich folglich um ein gemeinsames kooperatives Vorhaben dieser vier Akademien, das in zwei Arbeitsstellen an der BBAW und der Göttinger Akademie umgesetzt wird. Darüber hinaus möchte ich keinesfalls die weiteren Kooperationspartner wie das Leibniz-Institut für Deutsche Sprache in Mannheim, heute vertreten durch seinen Direktor Henning Lobin, die Deutsche Akademie für Sprache und Dichtung in Darmstadt und das Goethe-Institut in München unerwähnt lassen. Den Präsidenten des Goethe-Instituts Klaus-Dieter Lehmann, der auch Mitglied der BBAW ist, möchte ich besonders begrüßen; er wird heute Abend den Festvortrag halten.

Wir freuen uns alle gemeinsam, dass das Bundesministerium für Bildung und Forschung den heutigen Startschuss für das ZDL ermöglicht hat, und sind sehr dankbar dafür.

Das Deutsche kann auf eine lange und reiche Tradition der Wörterbucharbeit zurückblicken, wie sie nur für wenige andere Kultursprachen besteht. Sie reicht

vom monumentalen *Deutschen Wörterbuch* der Brüder Grimm über die unterschiedlichsten Spezialwörterbücher bis zu zweisprachigen Wörterbüchern für Reisende.

Erlauben Sie mir, an dieser Stelle einen kleinen historischen Überblick über die Lexikographie zu geben; ich werde dabei nur einige wenige ausgewählte Meilensteine nennen. Die Anfänge der deutschen Lexikographie liegen in mittelalterlichen Glossarien. Ein zweisprachiges Werk in Latein und Althochdeutsch, der *Codex Abrogans* aus dem 8. Jh., gilt als das älteste erhaltene Buch in deutscher Sprache. Solche Glossarien wurden für den Sprachunterricht genutzt, aber auch um Reisenden die Kommunikation zu erleichtern. Bis zum 17. Jahrhundert entstand eine Reihe von weiteren wichtigen Werken der deutschen Lexikographie, wie beispielsweise das *Onomasticon latinogermanicum* oder *Nomenclator Latinosaxonicus*. Weitere Meilensteine der deutschen Lexikographie sind Georg Henischs Wörterbuch *Teütsche Sprach und Weissheit: Thesaurus lingvae et sapientiae Germanicae* aus dem 17. Jahrhundert und Matthias Kramers *Das herrlich Grosse Teutsch-Italiänische Dictionarium oder Wort- und Red-Arten-Schatz* vom Beginn des 18. Jahrhunderts.

Das erste einsprachige Wörterbuch der deutschen Sprache stammt von Johann Christoph Adelung vom Ende des 18. Jahrhunderts. Es trägt den umfangreichen Titel *Versuch eines vollständigen Grammatisch-kritischen Wörterbuchs der Hochdeutschen Mundart mit beständiger Vergleichung der übrigen Mundarten, besonders aber der oberdeutschen* und umfasst mehr als 55.000 Einträge in strikt alphabetischer Reihenfolge. Diese überragende Leistung eines einzelnen Wissenschaftlers fand nicht zuletzt durch seine Nutzerfreundlichkeit eine weite Verbreitung und hat somit auch maßgeblich zur Etablierung und Ausbreitung der deutschen Standardsprache beigetragen.

Einen wahren Meilenstein der Lexikographie – auch im internationalen Maßstab – stellt das *Deutsche Wörterbuch* (DWB) der Brüder Jacob und Wilhelm Grimm, kurz *der Grimms*, aus dem 19. Jahrhundert dar. Die Grimms begannen mit ihrer Arbeit am Wörterbuch im Jahr 1838 und glaubten nach 10 Jahren fertig zu sein. Am 20. September 1863, dem Tag seines Todes, also 25 Jahre nach Beginn, arbeitete Jacob Grimm am Wort „Frucht“. Die Grimms sind trotz immenser Arbeitsleistung nur bis zum Buchstaben „F“ gekommen. Die Last der Fortführung dieses gewaltigen Werks wurde von vielen Lexikographen getragen. Im Jahr 1908 übernahm die BBAW-Vorgängerin, die Preußische Akademie der Wissenschaften, das Projekt. Der

letzte Band des Wörterbuches wurde jedoch erst im Januar 1961 – also 123 Jahre nach Planungsbeginn – in Kooperation der Berliner und der Göttinger Akademie der Wissenschaften veröffentlicht. In den 1950er Jahren wurde bereits mit der Überarbeitung des damals schon überholten Grimm begonnen.

Aus der kritischen Beschäftigung mit dem *Grimm'schen Wörterbuch* heraus entstanden mehrere Wörterbücher, so u. a. das *Wörterbuch der deutschen Sprache, mit Belegen von Luther bis auf die Gegenwart*. Dieses stammt von Daniel Sanders, der nach Ansicht vieler neben den Brüdern Grimm als bedeutendster deutscher Lexikograph des 19. Jahrhunderts gilt und dessen Geburtstag sich in diesem Jahr zum zweihundertsten Male jährt.

Im Jahr 1952 begann in der DDR an der Deutschen Akademie der Wissenschaften zu Berlin unter der Herausgeberschaft von Ruth Klappenbach und Wolfgang Steinitz ein einsprachiges Wörterbuchprojekt, das sich mit der deutschen Gegenwartssprache befasste. Die Einträge dieses *Wörterbuches der deutschen Gegenwartssprache* (WDG) beruhen auf sehr sorgfältig zusammengestellten Primärquellen und zeichnen sich durch überaus klare Bedeutungsdefinitionen aus. In der Bundesrepublik entstanden etwas später zwei vergleichbare mehrbändige Werke – Günther Drosdowskis *Duden: Das große Wörterbuch der deutschen Sprache in sechs Bänden!* und der *BrockhausWahrig: Deutsches Wörterbuch in sechs Bänden*. Der einbändige *Duden: Deutsches Universalwörterbuch*, dessen erste Auflage aus dem Jahr 1983 stammt, hat sich als Standard-Wörterbuch der deutschen Sprache etabliert.

In den letzten Jahrzehnten wurden viele wissenschaftliche Wörterbücher unter aktuellen wissenschaftlichen Aspekten überarbeitet oder neu aufgelegt – zum Teil unter Zuhilfenahme neuer technischer Errungenschaften. So ist beispielsweise an unserer Akademie das *Digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts* (DWDS) entstanden, das auf einer Sammlung von Korpora basiert, die insgesamt über 13 Milliarden Wörter umfassen und dessen Kernkorpus frei zugänglich im Internet online recherchiert werden kann. Darüber werden wir nachher noch mehr hören, denn auf ihm beruht die technische Plattform des ZDL. Sie sehen also, dass das ZDL in große Fußstapfen tritt. Das ängstigt uns nicht, sondern spornt uns an. Wenn uns das, was wir vorhaben, gelingt, und ich gehe davon aus, dass das so sein wird, werden wir mit dem ZDL, den in diesem Projekt zu entwickelnden algorithmischen Werkzeugen und den sich daraus ergebenden

Möglichkeiten der Analyse der deutschen Sprache und ihrer Entwicklung künftig mit an der Weltspitze der Lexikographie stehen.

Lassen Sie mich zum Schluss ein Geständnis machen. Davon wissen nur eine Handvoll Personen, aber manchmal ist es angebracht, Fehler einzugestehen. Mitte der 80er Jahre war ein mit einem Humboldt-Forschungspreis ausgezeichnete US-amerikanische Mathematiker für ein Jahr mein Gast. Er und seine Frau nahmen sich vor, in dieser Zeit Deutsch zu lernen. Ich sagte damals zu den beiden. „Ich helfe euch gerne. Und wenn ihr einmal ein Wort nicht kennt, erkläre ich es euch. Ich denke, ich kenne jedes Wort der deutschen Sprache.“ Das war, wie ich heute weiß, die größte Fehleinschätzung meines Lebens. Ich habe daraus gelernt. (Ich verrate Ihnen jetzt nicht, wie viele Wörter die deutsche Sprache nach heutigen Kenntnissen hat. Das überlasse ich nachfolgenden Rednern.) Mein amerikanischer Kollege begann sein Deutschstudium mit dem Lesen von Grimms Märchen. Sehr bald stieß er auf das Märchen mit dem Titel *Der Gaudieb und sein Meister*. Auf seine Frage, was denn ein „Gaudieb“ sei, musste ich nach intensiver Durchsicherung meines Gedächtnisses beschämt antworten, dass ich das Wort noch nie vorher gehört hatte. Mit dieser Beichte gestehe ich jetzt die Sünde der Hybris ein und hoffe auf Absolution durch meinen persönlichen Einsatz für das ZDL.

Meine sehr verehrten Damen und Herren, liebe Kolleginnen und Kollegen, mit dieser kleinen Anekdote möchte ich schließen. Ich wünsche uns allen einen anregenden weiteren Verlauf des Tages und bin sicher, dass er uns viele neue Einblicke in die Lexikographie bescheren wird.

**Andreas Gardt** ist Präsident der Akademie der Wissenschaften zu Göttingen und Mitglied der Deutschen Akademie für Sprache und Dichtung. Er hat die Professur für Germanistische Sprachwissenschaft und Sprachgeschichte an der Universität Kassel inne und forscht zur Geschichte der Sprachtheorie, zum Beitrag von Sprache zur Bildung kultureller und politischer Identität sowie zur Theorie und Praxis der Analyse von Texten und gesellschaftlichen Diskursen. Andreas Gardt ist wissenschaftlicher Leiter des Projekts „Wortgeschichte digital“, das im Rahmen des ZDL an der Göttinger Akademie der Wissenschaften erarbeitet wird.



Foto: Andreas, Fischer

## GRÜßWORT

Ich spreche in dieser Veranstaltung in zweifacher Funktion: als Präsident der Akademie der Wissenschaften zu Göttingen und als ein Wissenschaftler, der persönlich im Zentrum für digitale Lexikographie der deutschen Sprache engagiert ist. Zugleich spreche ich für die Sächsische und für die Mainzer Akademie, die unser Vorhaben mittragen. An der Sächsischen Akademie wird an einem Althochdeutschen Wörterbuch gearbeitet, an der Mainzer an einem Mittelhochdeutschen, gemeinsam mit der Göttinger Akademie, an der auch noch ein Frühneuhochdeutsches Wörterbuch entsteht. Die Wörterbuchlandschaft in Deutschland ist, jedenfalls wo es um die deutsche Sprache geht, vielfältig und die Akademien kooperieren intensiv.

In Göttingen begann am 1. Januar dieses Jahres die Arbeit an „Wortgeschichte digital“, unserem Teilprojekt des ZDL. Volker Harm, der Arbeitsstellenleiter, wird von unserem Vorhaben berichten. Ausdrücklich danke ich an dieser Stelle Wolfgang Klein und Martin Grötschel für ihr Engagement für das ZDL, das entscheidend für das Gelingen des Vorhabens war.

Bislang nicht erwähnt habe ich das Wörterbuch, das unter den genannten dem ZDL am nächsten steht: das *Deutsche Wörterbuch*, begründet von Jacob und Wilhelm Grimm. Sein erster Band erschien 1854, die letzte Lieferung 1961, der letzte Band der Neubearbeitung der Strecke A bis F, die in den Händen der Berlin-Brandenburgischen und der Göttinger Akademie lag, erst vor wenigen Monaten.

Das *Grimm'sche Wörterbuch* ist so etwas wie einer Kathedrale der Lexikographie, und das nicht nur im Hinblick auf seinen schiereren Umfang. Man halte sich das Ende von Jacob Grimms Vorwort zum ersten Band vor Augen:

„Deutsche geliebte landsleute, welches reichs, welches glaubens ihr seiet, tretet ein in die euch allen aufgethane halle eurer angestammten, uralten sprache, lernet und heiliget sie und haltet an ihr, eure volkskraft und dauer hängt in ihr.“

So würde man heute kein Wörterbuch mehr einleiten, jedenfalls nicht bei uns, und das aus gutem Grund. Aber man soll sich nicht täuschen lassen: Nationales Pathos und bürgerliches Freiheitsdenken gingen bei den Grimms Hand in Hand. Wenn bei ihnen, vor dem Hintergrund etwa von Jacobs Präsenz in der Frankfurter Paulskirche, von „volkskraft“ die Rede ist, dann ist das nicht dasselbe, was in Texten anderer, vor allem zwei oder drei Generationen später, damit verbunden war.

Damit ist man mit einem Schlag bei einer lexikographischen Selbstverständlichkeit angelangt: dass sich die Bedeutungen von Wörtern ändern. Und dass dies immer, in jedem einzelnen Fall, etwas mit denjenigen zu tun hat, die sprechen oder schreiben. Kein Werk, das der systematischen Erfassung von Sprache dient, erlaubt einen solch direkten Zugriff auf das Wissen, Denken und Fühlen der Menschen wie ein Wörterbuch. Keine Grammatik kann das leisten, auch wenn Grammatisches natürlich auch an semantischen Veränderungen beteiligt ist, man denke nur an die aktuelle Verschiebung von „Flüchtling“ zu „Geflüchteten“, doch auch dabei ist nur die Semantik der Grammatik wichtig, weil sie die Bedeutung nuanciert. Die Beschäftigung mit Fragen der Bedeutung, mit ihrer Gegenwart vor dem Hintergrund ihrer Geschichte, wird dem Ort und der Aufgabe von Sprache in der Welt in ganz besonderem Maße gerecht.

Lexikographen wissen um die Relativität von Bedeutung, und sie wissen auch, dass die Wörter nicht nur die Welt der Sprecher spiegeln, sondern ihr Denken und Fühlen prägen. In sprachphilosophischen Texten wird seit Jahrhunderten die

Trias von Sprache, Denken und Wirklichkeit diskutiert, und nicht erst in modernen konstruktivistischen Zeiten wird dabei der Sprache meist das Apriori zuerkannt. Man mag einwenden, die sprachliche Prägung von Denken und Welt geschehe nicht auf der Ebene einzelner Wörter, sondern durch Texte und Gespräche, die wiederum Teile gesellschaftlicher Diskurse sind. Das ist richtig, aber die Wörter sind die semantischen Bausteine von Texten und Diskursen. Man muss sich nur vor Augen halten, worüber wir in der Debatte über die Political Correctness streiten: ob wir dieses Wort sagen oder jenes meiden sollten.

Aber auch jenseits dieser Debatte gilt das. Um noch einmal auf „Volk“ zurückzukommen, wie es im Brockhaus Konversations-Lexikon definiert wird (zwar eine Enzyklopädie, kein Wörterbuch, aber im ersten definitiven Zugriff einem Wörterbuch vergleichbar):

„Volk“ 1908: „ein natürliches Ganzes bildende[r] Teil der Menschheit“  
„Volk“ 1994: „eine Gruppe von Menschen, die sich als ideelle Einheit begreift“

– damals natürliche, heute selbst gesetzte Größe (was auch den Ruf „Wir sind das Volk“ erst plausibel macht).

Und es ist wahrhaftig ein großer Unterschied, ob man „Gott“ als „höchstes übernatürliches Wesen“ definiert (wie der große Duden es um das Jahr 2000 tut) oder ob man ihn (wie es nahezu zeitgleich im Wörterbuch Deutsch als Fremdsprache [2000] geschieht) „als überirdisch und allmächtig gedachtes ... Wesen“ (Hervorhebung A. G.) beschreibt.

Ermöglicht durch die Förderung des Bundesministeriums für Bildung und Forschung und unter dem Dach der Union der deutschen Akademien der Wissenschaften haben wir mit dem ZDL eine ebenso wichtige wie attraktive Aufgabe übernommen. Das Ergebnis unserer Arbeit wird allen zugutekommen, die sich für die deutsche Sprache interessieren.

**Thomas Rachel** ist Parlamentarischer Staatssekretär bei der Bundesministerin für Bildung und Forschung. Seine inhaltlichen Schwerpunkte im Bundesministerium sind das Wissenschaftssystem, die Hochschulen und die Forschungsorganisationen, die Lebenswissenschaften und Forschung für die Gesundheit, die europäische und internationale Zusammenarbeit in Bildung und Forschung sowie der Haushalt des Ministeriums.



Foto: Bundesregierung, Sandra Stein

## GRUßWORT

Sehr geehrte Präsidenten, Herr Professor Grötschel und Herr Professor Gardt, sehr geehrter Herr Staatssekretär Krach, sehr geehrte Damen und Herren,

kaum ein Thema hat die Wissenschaft, Wirtschaft und Politik in diesem Jahrzehnt so beschäftigt wie die Digitalisierung. Mit enormer Wucht dringt sie in so viele Lebensbereiche, ruft einerseits Aufbruchstimmung, andererseits Skepsis hervor. Die Digitalisierung begeistert, sie polarisiert, sie stellt vieles auf den Kopf.

Das hat auch die starke Tradition der deutschen Wörterbücher erfahren müssen. Wer früher wissen wollte, wie ein Wort geschrieben oder benutzt wird, der griff ins Bücherregal und schaute beispielsweise im *Duden* nach. Der *Duden* war eine Instanz der deutschen Sprache. Wurde ein neues Wort aufgenommen, konnte diese Entscheidung die Feuilletons füllen und die Gemüter erhitzen.

Wie anders sieht das heute aus: Man zückt das Smartphone, binnen Sekunden findet man eine Fülle von Angeboten. Die Frage, warum ein Wort aufgeführt wird, stellt sich nicht mehr, sondern eher die Frage, warum ein Wort nicht aufgeführt wird –

ist man mittlerweile doch gewohnt, im Internet alles zu finden, und vieles vermeintlich ohne Gegenleistung.

Die zahlreichen kostenlosen Nachschlagemöglichkeiten im Internet sind selbst für die großen Wörterbuchverlage zum Problem geworden. So wurde z. B. das andere große deutsche Wörterbuch neben dem *Duden*, der *Wahrig*, 2014 eingestellt.

Ich denke, vielen Menschen ist klar: Wissenschaftliche Verlässlichkeit kann man von solchen Angeboten nicht erwarten. Aber Sprache ist eine mächtige Lenkerin, die Denken, Empfinden und Wertvorstellungen prägen kann. Mit der Sprache können sich auch Sichtweisen verschieben in einem positiven Sinn genau wie in einem negativen Sinn. Ein wissenschaftlich fundierter Umgang mit unserer Sprache stellt in meinen Augen deshalb ein gesamtgesellschaftliches Anliegen mit zukunftsweisender Verantwortung dar.

Die Digitalisierung und das Internet müssen nicht das Ende der wissenschaftlichen Lexikographie bedeuten. Sie bedeuten eine grundlegende Veränderung, und diese Veränderung sehe ich durchaus positiv.

Wörterbucharbeit kommt praktisch nie zu einem Ende. Noch während man am Buchstaben B arbeitet, könnte man schon den Buchstaben A korrigieren, ergänzen und aktualisieren. Bei den Erscheinungszyklen gedruckter Ausgaben war das nicht möglich. Ein digitales Wörterbuch aber kann man jederzeit korrigieren, ergänzen und aktualisieren. Für die Lexikographie stellt dies eine großartige Chance dar.

Die Akademien, die über Jahrzehnte an großen Wörterbuchprojekten gearbeitet haben, sehen diese Chance. Und sie möchten sie mit dem Zentrum für digitale Lexikographie der deutschen Sprache nutzen. Mit dem ZDL wird die lange und erfolgreiche Tradition der deutschen Wörterbucharbeit in das digitale Zeitalter überführt.

Aus Sicht des Bundesministerium für Bildung und Forschung (BMBF) ist dies nicht nur der Bedeutung des Deutschen als einer Kultursprache angemessen, es ist auch ein notwendiger Schritt, um die internationale Anschluss- und Wettbewerbsfähigkeit der deutschen Lexikographie, Germanistik, Sprachwissenschaft und weiterer benachbarter Fächer zu erhalten und auszubauen. Das BMBF unterstützt den Aufbau des Zentrums für digitale Lexikographie deshalb mit einer zunächst

fünffährigen Projektförderung in Höhe von fast 11 Mio. Euro. Bei erfolgreicher Evaluierung können drei weitere Jahre mit 6 Mio. Euro folgen.

Je weiter das Digitale um sich greift, desto offenkundiger wird, dass es eben nicht kostenlos zu haben ist, sondern ganz eigene Kosten damit verbunden sind. Server müssen nicht nur angeschafft, sondern auch unterhalten werden. Elektronische Bearbeitungswerkzeuge müssen nicht nur entwickelt, sondern auch unterhalten werden. Es müssen Personen und Tätigkeiten dauerhaft finanziert werden, wenn Pilotprojekte längst abgeschlossen sind.

Die Politik hat dies mittlerweile erkannt. Die Gemeinsame Wissenschaftskonferenz von Bund und Ländern hat deshalb die NFDI – die Nationale Forschungsdateninfrastruktur auf den Weg gebracht. Die NFDI soll eine koordinierte, vernetzte Informationsinfrastruktur werden, die ein übergreifendes Forschungsdatenmanagement und die Steigerung der Effizienz des gesamten Wissenschafts-systems verfolgt. Dazu soll sie ein verlässliches und nachhaltiges Dienste-Angebot schaffen, welches übergreifende und fachspezifische Bedarfe des Forschungsdatenmanagements in Deutschland abdeckt. Dies hat die Gemeinsame Wissenschaftskonferenz Ende vergangenen Jahres beschlossen. Ein von der DFG organisierter Wettbewerb wird jetzt geeignete Konsortien identifizieren, um langfristige Strukturen aufzubauen.

Die Möglichkeiten der BMBF-Projektförderung sind zeitlich begrenzt. Sie kann nur den Aufbau von Strukturen anstoßen, die Verstetigung muss anders erreicht werden.

Das BMBF hat mit dieser Art der Strukturbildung in den vergangenen Jahren gute Erfolge erzielt, weil die Verstetigung frühzeitig mitgeplant wurde. Beispiele dafür sind die eHumanities-Zentren für die digitalen Geisteswissenschaften wie das „Kölner Zentrum Analyse und Archivierung audiovisueller Daten“ oder das „Zentrum Musik – Edition – Medien“ an der Universität Paderborn. Aufgabe dieser Zentren ist es, Wissenschaftlerinnen und Wissenschaftlern elektronische Ressourcen, Werkzeuge und Methodenkenntnisse für ihren individuellen Forschungsprozess zur Verfügung zu stellen. Diese eHumanities-Zentren werden dauerhaft bestehen, auch wenn die Projektförderung ausläuft.

Meine sehr geehrten Damen und Herren, ich hatte es gerade schon angesprochen: Aus Sicht des BMBF ist das Zentrum für digitale Lexikographie der

deutschen Sprache eine wichtige Einrichtung, um die internationale Anschluss- und Wettbewerbsfähigkeit der deutschen Lexikographie, Germanistik, Sprachwissenschaft und weiterer benachbarter Fächer zu erhalten und auszubauen.

So komplex das Spannungsfeld von Selbstzweck und Nutzen von Wissenschaft ist, wird sich das ZDL genau wie die Angebote der Nationalen Forschungsdateninfrastruktur daran messen lassen müssen, wie sehr es genutzt wird. Meine Hoffnung ist es deshalb, dass Sie alle dazu beitragen, dass es ein nützliches Zentrum wird. Sie können dazu beitragen, indem Sie Nutzerinnen und Nutzer werden, aber auch indem Sie sich an den Veranstaltungen, den Workshops und der Weiterentwicklung des Zentrums beteiligen. Der heutige Tag ist dafür der Auftakt.

Ich wünsche dem Zentrum für digitale Lexikographie der deutschen Sprache viel Erfolg und uns allen heute einen interessanten Nachmittag.

**Steffen Krach** ist im Berliner Senat seit 2014 Staatssekretär für Wissenschaft und seit 2016 für Wissenschaft und Forschung. Zwischen 2007 und 2012 hatte er verschiedene Positionen in der Senatsverwaltung für Bildung, Wissenschaft und Forschung inne. 2012 wechselte Steffen Krach auf die Bundesebene und leitete bis 2014 die Bund-Länder-Koordinierungsstelle der SPD-Bundestagsfraktion.



Foto: Senatskanzlei, Lukas Hofmann

## GRUßWORT

Sehr geehrter Herr Grötschel, sehr geehrter Herr Gardt, sehr geehrter Herr Rachel, sehr geehrte Abgeordnete aus dem Deutschen Bundestag, meine sehr verehrten Damen und Herren,

ich möchte Sie alle herzlich begrüßen und Ihnen die herzlichsten Grüße vom Regierenden Bürgermeister ausrichten, der im Land Berlin nicht nur Regierender Bürgermeister, sondern auch Senator für Wissenschaft und Forschung ist. Der Regierende Bürgermeister hätte heute gerne teilgenommen, er konnte aber leider nicht kommen, weil wir zeitgleich eine gemeinsame Kabinettsitzung der Länder Berlin und Brandenburg haben.

Lassen Sie mich zunächst vor allem darauf hinweisen, dass ein so großes und auf lange Sicht angelegtes Vorhaben wie das Zentrum für digitale Lexikographie der deutschen Sprache nur unter Beteiligung mehrerer Akteure möglich ist. In diesem Fall sind das die Berlin-Brandenburgische Akademie der Wissenschaften, die Göttinger Akademie der Wissenschaften, die Mainzer Akademie der Wissenschaften und Literatur sowie die Sächsische Akademie der Wissenschaften. Gerade die Zusammenarbeit dieser Institutionen, die unter der Schirmherrschaft der Union der Deutschen Akademie

der Wissenschaften steht, stellt einen ganz besonderen Wert dar, der Qualität und Kontinuität garantiert. Das Gleiche gilt für die ZDL-Kooperationspartner wie zum Beispiel das Leibniz-Institut für Deutsche Sprache in Mannheim.

Ich freue mich darüber, dass das Bundesministerium für Bildung und Forschung eine Verstärkung des ZDL in Aussicht gestellt hat, die nach der Anlaufphase – natürlich nur nach Vorlage eines überzeugenden Konzepts für den Dauerbetrieb und einer erfolgreichen Evaluierung – erfolgen kann. Es ist auch schön zu sehen, dass die Erforschung des Wortschatzes der deutschen Sprache ein gesamtdeutsches Vorhaben ist.

Ich möchte mich ganz herzlich dafür beim BMBF bedanken, dass hier ein finanzielles Engagement am Standort Berlin vorgenommen wird. Ich möchte dies betonen, denn das Engagement nicht nur in Berlin, sondern in vielen anderen Bundesländern im Bereich von Wissenschaft und Forschung durch das BMBF ist essenziell und hilft den Vertreterinnen und Vertretern der Länder, die Strukturen in Wissenschaft und Forschung aufrechtzuerhalten und auszubauen.

Die deutsche Sprache ist fraglos eine bedeutende Wissenschafts- und Kultursprache. An ihrer Untersuchung und den dabei gewonnenen Einsichten besteht naturgemäß nicht nur ein wissenschaftliches, sondern auch ein großes gesellschaftliches Interesse. Mithilfe des Zentrums für digitale Lexikographie der deutschen Sprache wird der deutsche Wortschatz in Geschichte und Gegenwart zum einen umfassend und verlässlich beschrieben und zum anderen allen Interessierten über eine digitale Plattform frei zugänglich gemacht. Das wird nicht nur für die Wissenschaft, sondern für alle, die die deutsche Sprache verwenden, von großer Bedeutung sein. Gestatten Sie mir einen kurzen Blick in die Historie aus Berliner Sicht. Bereits bei der Gründung der Akademie im Jahr 1700 – damals hieß sie Kurfürstlich-Brandenburgische Sozietät der Wissenschaften – betonte der brandenburgische Kurfürst Friedrich III., dass „man auch auf die Kultur der deutschen Sprache bei dieser Fundation gedenken möchte“. Dem Kurfürsten schwebte ein umfangreiches Wörterbuch der deutschen Sprache vor, mit dem eine einheitliche Nationalsprache dokumentiert und verfestigt werden sollte. Mit der Akademiegründung wurde der Akademie somit nicht nur eine wichtige wissenschaftliche Aufgabe übertragen, der Kurfürst verband mit der lexikographischen Arbeit zugleich auch eine politische Dimension.

Die Akademie hat den Wunsch des Kurfürsten in dessen Lebenszeit nicht erfüllen können, hat sich aber in den 320 Jahren seit ihrer Gründung auf vielfältige Weise

mit der deutschen und mit anderen Sprachen befasst. Davon werden Sie heute noch mehr in weiteren Vorträgen hören.

Vor diesem Hintergrund könnte man fast sagen, dass die Akademie den Auftrag ihres kurfürstlichen Gründers nun wieder aufnimmt und dass das Zentrum für digitale Lexikographie der deutschen Sprache, dessen Hauptarbeitsstelle an der BBAW angesiedelt ist, eine Weiterführung der politischen Planung von 1700 darstellt. In jedem Fall ist durch das ZDL Berlin – zusammen mit Göttingen, wo an der dortigen Akademie, die ebenfalls eine große lexikographische Historie aufzuweisen hat, die zweite Arbeitsstelle des ZDL eingerichtet wird – die Lexikographie nach über dreihundert Jahren wieder ganz nach vorn gerückt.

Dies ist für die Germanistik in Deutschland im Allgemeinen und speziell im Hinblick auf den Wissenschaftsstandort Berlin für die geisteswissenschaftliche Forschung eine sehr erfreuliche Entwicklung.

Der Wissenschaftsstandort und Forschungsstandort Berlin steht glücklicherweise auf einem sehr breiten Fundament und ist in den Geisteswissenschaften ebenso attraktiv wie in den Naturwissenschaften, den Ingenieurwissenschaften oder auch in den Lebenswissenschaften. Gerade dieser besondere „Berliner Mix“, der in der Berlin-Brandenburgischen Akademie der Wissenschaften wie an kaum einem anderen Ort in der Stadt erlebbar wird, ist ein gutes Rüstzeug, um den Herausforderungen der Gegenwart und den technologischen Entwicklungen mit Klugheit zu begegnen.

Herr Grötschel hat es im Vorgespräch gesagt: Es ist heute schon sehr umfangreich in den Medien über diese Eröffnung berichtet worden, unter anderem auch in der *Berliner Morgenpost*. Ein Satz hat mich als Staatssekretär für Wissenschaft und Forschung ganz besonders erfreut. Im Artikel stand: „Deshalb ist ein Projekt, wie wir es jetzt angehen, nirgendwo anders als in Berlin möglich.“ Ich könnte dem hinzufügen: Das gilt auch noch für viele weitere Projekte, die wir gemeinsam vorhaben. Aber ich will es dabei erst einmal belassen. Ich wünsche Ihnen allen eine spannende Veranstaltung. Ich wünsche dem Zentrum alles Gute für die Zukunft. Ich bin mir ganz sicher, Sie werden erfolgreich sein. Ich möchte mich den Wünschen von Herrn Rachel anschließen, dass es natürlich auch darum geht, einen möglichst breiten Zugang zu ermöglichen, der zu einer umfangreichen Nutzung einlädt, und dass wir in einigen Jahren sagen können, das Engagement des Bundes hat sich gelohnt, und wir werden dieses Engagement weiter gemeinsam verstetigen. Alles Gute.

**Wolfgang Klein** ist wissenschaftlicher Koordinator des Zentrums für digitale Lexikographie der deutschen Sprache, Leiter des Digitalen Wörterbuchs der deutschen Sprache und des Zentrums Sprache an der BBAW. Er war Professor für Sprachwissenschaften in Heidelberg und Frankfurt am Main und Direktor am Max-Planck-Institut für Psycholinguistik in Nijmegen. Seine Forschung konzentriert sich auf die Bereiche Textlinguistik, Spracherwerb, Ausdruck von Raum und Zeit in der Sprache und linguistische Poetik. Er ist Mitbegründer der Zeitschrift für Literaturwissenschaft und Linguistik und Träger des Gottfried-Wilhelm-Leibniz-Preises der DFG.



Foto: BBAW, Judith Affolter

## GEGENWART UND ZUKUNFT DER DEUTSCHEN LEXIKOGRAPHIE<sup>1</sup>

Wozu, meine sehr verehrten Damen und Herren, wozu macht man Wörterbücher? Der amerikanische Arzt Oliver Wendell Holmes, Zeitgenosse der Brüder Grimm, noch heute für seine medizinischen Leistungen, seinerzeit aber auch als Dichter berühmt, hat einmal gesagt: „When I feel inclined to read poetry, I take down my dictionary. The poetry of words is quite as beautiful as that of sentences.“ Das ist ein eher ungewöhnlicher Gebrauch von Wörterbüchern, an den er da gedacht hat. Ich weiß auch nicht, ob er recht hat. Vielleicht können wir uns ja alle mal versprechen, heute Abend aus dem Schrank ein Wörterbuch herauszunehmen, falls Sie noch eines haben, und Holmes' Ansicht auf ihren Wahrheitsgehalt zu überprüfen. Sie werden vielleicht ganz unterschiedlich urteilen. Aber wie auch immer Ihr Urteil ausfällt, es war bestimmt nicht dies, was dem von einem meiner Vorredner schon erwähnten Kurfürsten Friedrich III. vor Augen gestanden hat, als er auf Leibnizens Vorschlag, eine Akademie in Preußen zu gründen, gesagt hat, „daß man auch auf die Kultur der deutschen Sprache bei dieser Fundation

1 Ich danke Stefanie Saier für die Transkription der Rede. Sie wurde hier von einigen Versprechern und sonstigen Errata bereinigt und etwas geglättet. Für ihre Hilfe dabei danke ich Julia Naji.

gedenken möchte, gleichwie in Frankreich eine eigene Akademie gestiftet“. Seine Vorstellung dabei war es, ein Wörterbuch zu schaffen wie das der Académie française, in der Absicht, für eine einheitliche Sprache in den preußischen Landen und vielleicht auch noch darüber hinaus zu sorgen, sodass sich die Untertanen untereinander, aber auch mit der Obrigkeit vernünftig und klar verständigen konnten. Nicht dasselbe, aber eng verwandt mit diesem Ziel ist das, was dann Leibniz darauf antwortete, nämlich, dass es dabei auch auf „die Erhaltung der deutschen Sprache in ihrer anständigen Reinigkeit“ ankäme. Es ist die Vorstellung, dass man die Sprache nicht sozusagen wild wachsen lassen darf wie einen Urwald, sondern dass man sie irgendwie hegen und pflegen müsse – eine Vorstellung, die ich selbst immer ein wenig bizarr gefunden habe, die aber heute noch eine sehr große Rolle spielt.

Vereinheitlichung und Erhaltung der Sprache sind zwei zentrale Ziele, derentwegen man Wörterbücher macht. Das gilt aber nicht für die ältesten Wörterbücher, die wir kennen und die schon vor gut 4.000 Jahren entstanden sind. Man hat im Zweistromland Tausende von zweisprachigen Tontäfelchen gefunden – lange Listen mit sehr vielen Wörtern. Auf der einen Seite war dann beispielsweise Akkadisch, auf der anderen Seite Sumerisch. Es waren elementare Wörterbücher, die wohl vor allen Dingen dem Handel gedient haben, vielleicht auch dem Tourismus – genau wissen wir das nicht. Das war das älteste Ziel, eigentlich eine Gruppe von Zielen, nämlich der praktische Nutzen, den man unmittelbar hat, den ganz normale Leute haben, die das Wörterbuch konsultieren. Das gibt es nach wie vor. Dies gilt natürlich besonders für zweisprachige Wörterbücher, aber durchaus auch für einsprachige. Dort kann man zum Beispiel nachsehen, wie man ein Wort richtig schreibt oder wie man es ausspricht. Man kann sich grammatische Informationen holen, also beispielsweise, ob es „gebiert“ oder „gebärt“ heißt. Das kann man in einem solchen Wörterbuch, beispielsweise dem *Duden*, nachschlagen. Man kann die Bedeutung von Wörtern ansehen, die man vielleicht ungefähr versteht oder auch nicht ungefähr versteht, über die es vielleicht einen Streit gibt – zum Beispiel das Wort „Hetzjagd“. Einsprachige Wörterbücher erfüllen also durchaus praktische Zwecke. Seit einigen Jahren gibt es noch einen weiteren praktischen Zweck, den man vorher nicht kannte: Wörterbücher spielen auch eine wichtige Rolle bei der Digitalisierung, in der Computerlinguistik und der maschinellen Sprachverarbeitung.

Damit habe ich drei wichtige Ziele genannt, zu denen man Wörterbücher macht: Standardisierung, Sprachpflege, praktische Zwecke verschiedener Art. Es gibt

noch ein viertes Ziel, an das man als Laie nicht primär denkt, Wissenschaftler aber wohl, wenn nicht gar in erster Linie. Das ist die Befriedigung der Neugier, oder wie der Wissenschaftler eher sagt, der Wissbegierde. Man will wissen, wie es eigentlich gewesen ist und auch noch jetzt ist. Als die Brüder Grimm mit ihrem großen Wörterbuch begonnen haben, hatten sie keine praktischen Zwecke und auch keines der beiden anderen Ziele im Sinn, sondern sie wollten die Geschichte der deutschen Sprache, so wie sie sich im Wortschatz darstellt, nachzeichnen, sodass jeder sie darin verfolgen kann. Das vierte Ziel ist es also, Wissen zu schaffen und zu dokumentieren.

Die vier Ziele stehen nicht in einem Widerspruch zueinander, obwohl sie sich in der Praxis in die Quere kommen können. In den vergangenen 200 Jahren hat man für das Deutsche eine Fülle von Wörterbüchern geschaffen, die diese Ziele in unterschiedlicher Form und unterschiedlicher Kombination realisiert haben. Vom kleinen Liliput-Wörterbuch, das die Älteren unter uns seinerzeit benutzt haben, um bei den Klassenarbeiten zu schummeln – heute verwendet man eher das Smartphone –, über die größeren einsprachigen Wörterbücher, Orthographie-Wörterbücher, Synonym-Wörterbücher, Aussprache-Wörterbücher usw. bis zu den monumentalen Wörterbüchern, wie sie von Akademien und Verlagen geschaffen worden sind. Wörterbücher gibt es natürlich auch für andere Sprachen. Aber unter den rund 7.000, die derzeit auf der Welt gesprochen und geschrieben werden, gibt es nur einige wenige, die einen dem Deutschen vergleichbaren Reichtum des Wortschatzes aufweisen oder die auf vergleichbare Weise lexikographisch aufgearbeitet und dargestellt worden sind – vielleicht zehn oder zwölf Sprachen insgesamt. Dies ist zurückzuführen auf Tätigkeiten zweier Instanzen: Das eine sind kommerzielle Verlage. Sie alle kennen solche Verlage – das Bibliographische Institut, das den Duden gemacht hat, Brockhaus, das lange Zeit den *Wahrig* gemacht hat, aber auch zum Beispiel Langenscheidt, Klett und andere kleinere oder sehr viel kleinere. Das ist die eine Seite. Auf der anderen Seite sind es die wissenschaftlichen Akademien, die die großen Akademien-Wörterbücher begonnen haben und zum Teil noch fortführen. Herr Grötschel hat diese „Epochenwörterbücher“, wie man sie auch nennt, bereits erwähnt. Ich will sie aber noch einmal nennen, weil sie für die Konzipierung und den Aufbau unseres Zentrums eine wesentliche Rolle spielen. Man hat nämlich unter den Akademien eine gewisse Abdeckung der ganzen Zeit, seit man von deutscher Sprache redet, also ungefähr seit den Tagen Karls des Großen, angestrebt. Es gibt zunächst das von der Sächsischen Akademie in Leipzig bearbeitete *Althochdeutsche Wörterbuch*, dessen Anfänge in das Jahr 1935 zurückreichen. Die erste Lieferung ist 1952 erschienen; die letzten bisher

vorliegenden Lieferungen sind den Buchstaben O bis R gewidmet. Daran schließt sich zeitlich das *Mittelhochdeutsche Wörterbuch* an, das Anfang der 90er Jahre begonnen wurde und gemeinsam von den Akademien in Mainz und in Göttingen bearbeitet wird; zuletzt erschienen ist der Buchstabe O. Darauf folgt zeitlich das *Frühneuhochdeutsche Wörterbuch*, das Anfang der 80er Jahre von Oskar Reichmann begründet und hauptsächlich auch bearbeitet wurde; es ist jetzt auch in der Obhut der Göttinger Akademie. Ich sehe Herrn Reichmann da drüben, ich sage jetzt hoffentlich nichts Falsches, aber nach meiner Schätzung sind noch ungefähr 20 Prozent zu bearbeiten. Herr Reichmann widerspricht nicht, die Größenordnung stimmt also. Diese drei und das *Grimm'sche Wörterbuch*, das den ganzen Zeitraum behandelt, seinen Schwerpunkt aber im Neuhochdeutschen hat, decken mit gewissen Überlappungen den ganzen Zeitraum der deutschen Sprachgeschichte ab. Gemeinsam mit dem, was die Verlage beigetragen haben, ist dies eine fantastische Tradition, ein fantastischer Stand der deutschen Lexikographie.

Aber wie gleichfalls von meinen Vorrednern schon angedeutet wird seit Ende des letzten Jahrtausends, also vor zwei Jahrzehnten, immer deutlicher, dass diese Tradition in eine Krise gerät. Die Verlage können die großen Wörterbücher überhaupt nicht mehr „stemmen“. Das liegt daran, dass sie immer weniger gekauft werden, und dies hat seine Ursache darin, dass man über das Internet so viel an Informationen frei bekommen kann, auch wenn man über die Qualität dieser Informationen sicher streiten kann. Deshalb haben die großen Verlage ihre beeindruckende Arbeit auf diesem Gebiet stark reduziert, in manchen Fällen sogar vollständig eingestellt. Das ist beklagenswert, aber es ist schlichtweg eine Tatsache. Für die vier Akademien-Wörterbücher tritt das Problem in dieser Form nicht auf, weil sie von der öffentlichen Hand, vom Staat, finanziert worden sind – aus gutem Grund. Dafür sind wir sehr dankbar. Aber auch da gibt es Probleme, die aus der außerordentlich langen Laufzeit herrühren. Die erste Lieferung des *Deutschen Wörterbuchs* der Brüder Grimm ist 1852 erschienen, die letzte 1961. Die Buchstaben A bis F haben die Brüder Grimm selbst noch bearbeitet, der letzte Artikel, den Jacob Grimm verfasst hat, wurde 1863 geschrieben. Ende der 50er Jahre des letzten Jahrhunderts haben die Akademien in Göttingen und Berlin gemeinsam beschlossen, zumindest die Artikel von A bis F, also jene, die auf die Grimms selber zurückgehen, neu zu bearbeiten. Der Buchstabe A ist 1962 hier in Berlin in Angriff genommen worden. Im Jahre 1965 ist die erste Lieferung erschienen, die letzte im Jahre 2004. 40 Jahre für den Buchstaben A. Man kann jetzt hochrechnen, wie lange die Neubearbeitung der ganzen Buchstabenstrecke von A bis Z dauern würde. Dabei kann man die Zeit für das A, also 40 Jahre, nicht einfach

mit 26 multiplizieren, weil die Strecke für manche Buchstaben halt sehr kurz ist: Mit X oder Y fangen nur wenige Wörter an. Aber wir können es ja einmal mit 20 multiplizieren. Dazu braucht man nicht einmal einen Mathematiker wie unseren Präsidenten: 40 Jahre mal 20, das sind 800 Jahre, die man bräuchte, wenn man in dieser Weise fortfahren würde. Vielleicht kann man es noch ein wenig beschleunigen, wenn man mehr Leute einstellt, aber das ist ungefähr die Größenordnung, mit der man rechnen muss. Das finanziert keine öffentliche Hand. Es wäre auch selbst dann völlig sinnlos, denn in 800 Jahren wird die deutsche Sprache eine ganz andere sein, wenn es sie denn überhaupt noch gibt. Das heißt mit anderen Worten, um das Jahr 2000 hat sich gezeigt und in den folgenden Jahren ist immer deutlicher geworden, dass das mit dieser klassischen Weise des Wörterbuchmachens, so verdienstvoll und so bedeutend und so sorgfältig sie ist, einfach nicht mehr geht. Sie ist nicht mehr finanzierbar, nicht mehr tragbar.

In einer solchen Situation kann man es entweder aufgeben, oder man kann sich vielleicht einmal ein wenig zurücklehnen und überlegen, was man eigentlich will. Und was man will, ist, den deutschen Wortschatz so umfassend und sorgfältig wie irgend möglich zu beschreiben. Was ist der deutsche Wortschatz? Er setzt sich aus Wörtern zusammen, aber es sind nicht nur einzelne Wörter, sondern sie stehen auch in gewissen Beziehungen, in gewissen Sinnzusammenhängen untereinander. Das also möchte man beschreiben. Dabei ergeben sich zwei Probleme. Das eine hängt mit dem einzelnen Wort zusammen, das andere mit dem Wortschatz insgesamt in seiner ganzen Fülle und in seinem ganzen Ausmaß. Das will ich hier ganz kurz an einigen Beispielen erläutern.

Was ist ein Wort? Ein Wort ist nicht etwas, was irgendwo auf dem Papier geschrieben steht oder was Sie auf der Leinwand da oben lesen können. Das, was man sinnlich wahrnehmen, also sehen oder hören kann, ist sozusagen die Art und Weise, wie ein Wort nach außen tritt und woran man es erkennen kann. Das eigentliche Wort ist immer ein Bündel von Eigenschaften, die irgendwie miteinander in Verbindung stehen und die in der Geschichte der Menschheit und auch heute noch in den meisten Sprachen überhaupt nicht geschrieben worden sind, sondern einfach nur in den Köpfen der Sprecher existierten – nicht eines einzelnen Sprechers, sondern sehr vieler Sprecher. Die Menschen in irgendeinem Gebirgstal in Neuguinea, wo die Leute niemals eine Schrift gehabt haben, haben auch eine Sprache, sie steht bloß nicht in einem Wörterbuch oder einer Grammatik, sondern sie steht nur im Gehirn der Menschen, die sie sprechen.

Wörter sind Bündel von Eigenschaften in den Köpfen der Sprecher. Was sind diese Eigenschaften? Man sollte hier zwischen primären und sekundären Eigenschaften unterscheiden. Erstere sind gleichsam die definierenden. Es muss zunächst eine wahrnehmbare Form geben. Das ist in aller Regel eine Folge von Lauten; in ungefähr der Hälfte der Sprachen kommt zu der Lautform noch eine Schriftform dazu, eine Folge von Buchstaben oder anderer Schriftzeichen. Man hat, wie die Linguisten sagen, eine phonetische Repräsentation und eine graphematische Repräsentation. Aber sie allein machen noch kein Wort aus. Mit der Lautfolge oder sonstigen Zeichenfolge muss sich eine Bedeutung verbinden, ein Bündel von semantischen Merkmalen. Das ist die dritte primäre Eigenschaftsgruppe. Das vierte Bündel umfasst jene Eigenschaften, die sich auf die Verbindung von Wörtern mit anderen Wörtern zur größeren Einheit beziehen: die grammatischen Eigenschaften. Das sind beispielsweise jene, nach denen die Wörter flektiert werden, die ihre Wortklasse bestimmen usw. Ein Wort ist eine Verbindung von phonologischen, graphematischen, semantischen und grammatischen Eigenschaften. Sie sind konstitutiv und in diesem Sinne primär, wobei man die Einschränkung machen muss, dass nicht alle Sprachen eine Schriftform haben. Wenn man also Wörter als Ganzes beschreiben will und nicht nur den Anspruch hat, irgendwie einen Teil davon zu beschreiben, dann muss man eigentlich all dieses beschreiben.

Sekundäre Eigenschaften sind nicht definierend, aber dennoch wichtig. Man möchte zum Beispiel wissen, was die Etymologie eines Wortes ist – wo kommt es eigentlich her, was ist der Ursprung. Viele Wörter, die wir heute im Deutschen benutzen, kommen aus dem Indogermanischen, andere sind aus anderen Sprachen übernommen. An vielen kann man es noch erkennen, an anderen gar nicht. Um ein berühmtes Beispiel anzuführen: Das Wort „Hängematte“, das Sie alle kennen, wahrscheinlich haben Sie die Sache auch schon benutzt, hat weder etwas mit Hängen noch mit Matte zu tun, sondern kommt von dem Wort „xamaq“ aus der Sprache der Arawak-Indianer, das über Spanisch „hamaca“ ins Deutsche übernommen worden ist, aber dann umgedeutet wurde; es hat also einen ganz merkwürdigen Ursprung. Man möchte wissen, wie sich die Wörter über die Geschichte hin entwickelt haben. Wann sind sie zum ersten Mal verwendet worden? Wann waren sie häufig, wann sind sie außer Gebrauch gekommen oder zumindest selten geworden? Es gibt sozusagen statistische Fakten in der Verwendung von einzelnen Wörtern. Es gibt Beziehungen zwischen Wörtern wie Synonymie usw. All dies sind keine definierenden Eigenschaften, aber es sind Dinge, die man auch gerne wissen möchte. Eine wirklich umfassende Beschreibung des deutschen Wortschatzes müsste eigentlich all dieses im Prinzip angeben können. Das ist

aufwendig und schwierig und nur in Grenzen machbar – aber man muss möglichst weit in diese Richtung gehen.

Jetzt will ich Ihnen das Gesagte, um es ein wenig plastischer zu machen, an zwei Wörtern, verbunden mit zwei Aufgaben, verdeutlichen. Wie würden Sie die Bedeutung des Wortes „grinsen“ beschreiben? [Pause] Was heißt „grinsen“? Wie unterscheidet es sich von zum Beispiel „lächeln“ oder „lachen“ oder „kichern“, die alle irgendwie miteinander verwandt sind? Ich würde gern hier einen kleinen Preis ausloben für die beste Definition. Sie können mal darüber nachdenken. Es ist sehr, sehr schwierig. Ich will es hier auch gar nicht versuchen. Und wenn Sie heute Abend, dem Vorbild Oliver Wendell Holmes folgend, einmal ein möglichst großes Wörterbuch in die Hand nehmen, dann wird es Sie amüsieren, die Definition der genannten Wörter nachzulesen. Man versteht als Deutschsprachiger das Wort „grinsen“ im Allgemeinen sehr leicht, Kinder können es. Seine Definition auch in den besten Wörterbüchern versteht man meistens nicht sehr gut.

Nun besteht der deutsche Wortschatz nicht nur aus Inhaltswörtern wie Verben, Nomina und dergleichen, sondern auch aus Funktionswörtern, die oft sehr gängig sind, zum Beispiel die Funktionswörter „schon“ und „noch“. Was bedeutet eigentlich „schon“? „Die Zuhörer schlafen schon.“ Oder „schlafen noch“. Sie alle kennen das Wort. Was bedeutet es? Ich sehe schon, dass einige hier nachdenken. Vor einigen Jahren habe ich einmal beim *Oxford English Dictionary* einen Vortrag gehalten und mir den Spaß gemacht, die Definitionen vorzulesen, die dieses beste aller Wörterbücher für die beiden Wörter „still“ und „yet“ gibt, und dann die versammelten Experten, Herausgeber, Vize-Herausgeber usw. gefragt, welche Wörter denn da wohl definiert werden. Keiner hat es gewusst, keiner im ganzen Saal. Nicht einmal geraten hat es jemand. Es ist außerordentlich schwierig, eine solche Definition zu geben, und dies bei Wörtern, die Kinder schon mit drei Jahren verwenden. Noch schwieriger wird es, wenn man sich überlegt, dass man nicht nur die Bedeutung dieser Wörter für sich beschreiben muss, sondern auch, was dann passiert, wenn man ein solches Wort in einen Satz einfügt. Nehmen wir den einfachen Satz „Goethe war schon krank“ oder, etwas drastischer „Goethe war schon tot“, was man spätestens ab April 1832 mit Recht sagen konnte. Nun kann man das Wort „schon“ auch an einer anderen Stelle einfügen, nämlich ganz am Anfang: „Schon war Goethe tot“. [Gelächter]. Das habe ich erwartet, dass hier nicht gegrinst, aber doch ein wenig gelacht wird. Man hat irgendwie die Vorstellung, das passiert sehr schnell, schneller jedenfalls als erwartet. Aber wieso kann man eigentlich nicht sagen oder zumindest sehr schlecht sagen: „Schon Goethe war tot“?

Ich überlasse das Ihnen hier, die Antwort zu finden. Sie finden sie nämlich in keinem Wörterbuch; das müsste aber eigentlich der Fall sein, wenn das Wörterbuch den Anspruch hat zu erfassen, was „schon“ zur Bedeutung eines Satzes beiträgt. All dies sind Probleme mit ganz einfachen Wörtern, genauer gesagt den drei Wörtern „grinsen“, „schon“ und „noch“. Aber das Deutsche hat dummerweise nicht nur drei oder vier Wörter, sondern es hat wie viele? Das weiß man nicht. Im Grunde kann man aus vorhandenen immer wieder neue Wörter bilden, eine Möglichkeit, die es in allen Sprachen gibt, die aber im Deutschen besonders stark ausgebildet ist. Wie viele werden aber tatsächlich gebraucht? Im *Grimm'schen Wörterbuch* sind ungefähr 350.000 Wörter beschrieben. Im englischen *Oxford English Dictionary*, das vom Grimm inspiriert wurde und in mancher Beziehung vergleichbar ist, sind es 600.000. Wir haben an der BBAW vor ein paar Jahren einmal untersucht, wie viele Wörter in einem Korpus der Länge von einer Milliarde Wörtern der Gegenwartssprache tatsächlich vorkommen. Sie dürfen raten. [Pause] Es sind ungefähr fünf Millionen. Bei dieser Zahl gibt es eine gewisse Fehlerquote, es können auch 4,7 oder 5,3 Millionen sein, aber die Größenordnung stimmt. In größeren Korpora würde man mit Sicherheit noch mehr Wörter finden. Sie alle beschreiben zu wollen, ist praktisch unmöglich. Auch hier möchte ich, wie vorhin schon, eine Rechnung machen, diesmal am Beispiel der vorhin erwähnten Neubearbeitung des Grimm. Stellen Sie sich einen Lexikographen vor, der ein Wort wie „schon“ oder „noch“ oder „grinsen“ oder sonst irgendeines beschreiben will und es nicht nur irgendwo anders abschreibt, sondern aus den Belegen nachprüft, ob sich vielleicht seine Bedeutung irgendwann verändert hat; oder es vielleicht gar überhaupt zum ersten Mal beschreibt. Wie lange würde er dafür brauchen? Sagen wir, er braucht im Schnitt zwei Stunden dafür. Das ist sehr kurz gerechnet, wie jeder weiß, der das schon einmal versucht hat. Dann schafft er im Jahr bei 220 Arbeitstagen 880 Wörter. In zehn Jahren 8.800. In hundert Jahren 88.000. Sie können jetzt weiterrechnen. Das heißt, es ist ganz sinnlos, in der klassischen Weise ein umfassendes Wörterbuch schreiben zu wollen, das den gegenwärtigen Wortschatz vollständig oder auch nur annähernd vollständig abdeckt. So groß kann ein Wörterbuch überhaupt nicht sein.

Man muss also anders vorgehen. Und hier wird das, was zu den großen Schwierigkeiten für die Verlage geführt hat, nämlich das Aufkommen digitaler Methoden, auf einmal zu einer Möglichkeit, das Problem auch zumindest in steter Annäherung zu lösen: Man muss gänzlich neue Methoden anwenden, Methoden, bei denen man zum Beispiel die Bearbeitung nicht von A bis Z macht, sondern sich zuerst einmal überlegt: Was sind denn die wichtigsten Wörter, was sind jene, die noch

nie beschrieben worden sind, was sind jene, deren Bedeutung sich erheblich geändert hat. Die kann man zunächst einmal aufnehmen oder halt revidieren und bearbeiten. Dann kann man schrittweise weitere Wörter ergänzen, und man kann zu den vorhandenen schrittweise immer weitere Informationen hinzufügen. Der Computer und die Digitalisierung machen genau dieses möglich.

Und das will ich jetzt ganz kurz ausführen. Man hat beim digitalen Vorgehen keine Platzprobleme wie bei einem traditionellen Wörterbuch. In der Neubearbeitung des *Grimm'schen Wörterbuchs* ist die Buchstabenstrecke A bis F ungefähr viermal so lang wie die entsprechende Strecke A bis F im alten Grimm, dessen rund 350.000 Wörter sich auf 32 Foliobände verteilen. Bei einer Neubearbeitung so vieler Wörter käme man bei gleicher Vorgehensweise also auf ungefähr 130 Foliobände, und man hätte damit vielleicht ein Zehntel des Wortschatzes von fünf Millionen insgesamt erfasst. Ein Computer und das Internet haben im Prinzip keinerlei Begrenzung im Umfang. Das ist ein erster Vorzug des digitalen Vorgehens. Ein zweiter ist: Man braucht nicht Wort für Wort in einer einmal gewählten Weise zu bearbeiten, sondern kann schrittweise oder, wie man heute gern sagt, modular vorgehen und dabei beispielsweise die Tiefe der Bearbeitung variieren. Man kann auch unterschiedliche Aufgaben auf verschiedene Forschungsstätten aufteilen, die dann gemeinsam an dem ganzen Werk mitwirken. Keine Umfangsbeschränkung, modulares Vorgehen. Ein dritter wichtiger Vorzug des digitalen Vorgehens ist, dass man die Beschreibung der Wörter nach all den Eigenschaften, die ich vorhin genannt habe, unmittelbar mit den Daten verbinden kann, auf denen die Beschreibung beruht. Früher hat man dazu Zettelkästen mit Belegen gesammelt, deren Umfang notgedrungen beschränkt war. Heute kann man sich riesige Korpora aufbauen, zu denen nicht nur der Lexikograph Zugang hat, sondern die man, soweit das denn rechtlich möglich ist, auch für jeden Nutzer freischalten kann. Das heißt, man kann nicht nur die Bedeutung und sonstige Eigenschaften, Schreibweise eines Wortes usw. angeben, sondern man kann den Betreffenden in die Lage versetzen, das sofort nachzuprüfen, wenn er es nicht glaubt. Der Benutzer kann sich auch, wenn er ein Wort unzulänglich beschrieben findet, die Verwendung in vielen Belegen anschauen und die Beschreibung für sich selber vertiefen oder korrigieren. Nicht zuletzt gilt dies für Wörter, die überhaupt noch nicht bearbeitet sind, ja, angesichts ihrer Anzahl auch nicht bearbeitbar sind. Ich sage noch einmal: fünf Millionen. Da kann man immerhin in den Korpora nachsehen und sich aus dem Zusammenhang rekonstruieren, was die Eigenschaften eines noch nicht beschriebenen Wortes sind. Man hat also einen nicht vollständigen, aber doch einen ersten Zugang zur Fülle des gesamten deutschen Wortschatzes.

Ein vierter Punkt, in dem digitale lexikalische Systeme dem klassischen gedruckten Wörterbuch überlegen sind, liegt in der Möglichkeit, verschiedene Darstellungsmethoden zu wählen. Man kann zum Beispiel die Bedeutung technischer Begriffe, sagen wir „Pleuelstange“, einfach durch ein Bild darstellen. Bei „grinsen“ könnte man einfach ein Video einfügen. Man kann die Aussprache durch professionelle Sprecher einsprechen lassen, wie das im *Digitalen Wörterbuch der deutschen Sprache* hier an der Akademie, aber auch beispielsweise im Online-Duden bereits geschieht. Für einen japanischen Deutschlerner, der sich das in Tokio übers Internet anhört, ist das natürlich unvergleichlich besser, als wenn man ihm eine Lautschrift vorlegt, bei der er sich, wie viele andere, nur annähernd vorstellen kann, wie das Wort wirklich klingt. Ein fünfter Vorzug schließlich – nicht der letzte, aber damit möchte ich aufhören – liegt darin, dass man alles schnell korrigieren kann. Anders als bei einem gedruckten Wörterbuch muss man nicht auf die nächste Ausgabe warten, was meist viele Jahre dauert. Beim digitalen Vorgehen kann man es praktisch sofort machen, wenn der Fehler entdeckt wird – sofern man das Personal und das Geld dafür hat.

Wir haben vor einigen Jahren hier an der BBAW begonnen, ein solches digitales lexikalisches System zu entwickeln, das in dieser Weise vorgeht, das schon erwähnte *Digitale Wörterbuch der deutschen Sprache*. Alexander Geyken wird gleich ausführlicher darauf eingehen, sodass ich mir das hier spare. Dieses System ist zwar noch beschränkt, aber es funktioniert nicht nur, sondern es wird bereits rege genutzt. Wir haben derzeit etwa vier Millionen Aufrufe im Monat und über 40.000 registrierte Benutzer aus aller Welt. Dieses System wird die Ausgangsbasis für das kommende Zentrum für digitale Lexikographie sein. Das ZDL ist aber nicht einfach ein Ausbau des bestehenden digitalen Wörterbuchs, sondern es schließt vieles andere ein. Darauf und wie es zustande gekommen ist, möchte ich jetzt kurz eingehen.

Im Jahre 2010 hat mich Karl Stackmann, der ganz wunderbare Altgermanist und damalige Leiter der Göttinger Wörterbuchkommission, angerufen und gesagt: „Wir müssen etwas tun. Der *Grimm* läuft in zwei Jahren aus, und die Neubearbeitung bei den anderen Wörterbüchern wird auch noch sehr lange dauern. Was kann man da machen?“ Wir haben uns dann ein paar Mal in Göttingen getroffen und gemeinsam einen Plan für ein nationales Wörterbuchinstitut in Deutschland konzipiert. Es sollte aus dem DWDS mit seiner digitalen Plattform mit den Epochenwörterbüchern bestehen. Dieser Plan hat sich damals aus allerlei Gründen nicht realisieren lassen. Aber als Martin Grötschel Präsident dieser

Akademie geworden ist, hat er in seiner typischen Art die Initiative ergriffen und mich gefragt, ob wir da etwas machen können. Ich habe dann unter Rückgriff auf die früheren Überlegungen ein kurzes Konzeptpapier geschrieben, mit dem wir zum Bundesministerium für Bildung und Forschung geeilt sind. Dort haben wir, möchte ich sagen, fast offene Türen eingerannt. Das mag auch damit zusammenhängen, dass die Sorge um die deutsche Sprache keine Angelegenheit einzelner Bundesländer ist, sondern, wenn das Wort erlaubt ist, eine nationale Aufgabe. Wir haben dann, nachdem das BMBF uns grünes Licht gegeben hat, mit einer Reihe anderer Einrichtungen, allen voran anderer Akademien, aber auch dem Leibniz-Institut für Deutsche Sprache und dem Goethe-Institut gesprochen, einen umfassenden Antrag geschrieben, mit vielen Kollegen diskutiert, und schließlich eingereicht, der nach ausführlicher Begutachtung im letzten November bewilligt wurde. Am ersten Januar dieses Jahres konnten wir mit der Arbeit beginnen.

Der Plan sieht vor, dass es in dem ganzen Vorhaben zwei Phasen gibt, zunächst eine Aufbauphase, die acht Jahre dauern soll, und eine Betriebsphase, die dann im Prinzip auf Dauer angelegt ist. Gefördert wird zunächst diese Aufbauphase, und zwar für einen der beiden Bereiche, die das ZDL ausmachen. Wir haben nämlich eine inhaltliche Unterteilung in Arbeiten zum Neuhochdeutschen ab ungefähr 1600 bis in die Gegenwart einerseits und für die drei früheren Sprachstufen Althochdeutsch, Mittelhochdeutsch und Frühneuhochdeutsch andererseits. Für letztere laufen ja noch die verschiedenen Akademien-Wörterbücher, die ich vorhin genannt habe, und sie sollen auch in ihrer bewährten Weise bis zu Ende geführt werden. Wie sie dann hinterher in das ZDL integriert werden, darüber muss während der Aufbauphase verhandelt werden. Für den Bereich des Neuhochdeutschen haben wir wiederum eine Unterteilung in eine Göttinger Arbeitsstelle und eine Berliner Arbeitsstelle. Die Göttinger Arbeitsstelle konzentriert sich darauf, in Form einer Anzahl von „Wortgeschichten“ bestimmte wichtige Wörter ausführlich im Zusammenhang zu anderen verwandten Wörtern zu beschreiben. Die Berliner Arbeitsstelle liefert den gesamten IT-Unterbau auf der Basis des DWDS. Sie wird weiterhin die Korpora, die wir bereits haben, in vielerlei Hinsicht erweitern, um die Fülle der deutschen Sprache abzudecken. Drittens schließlich wird sie den Bestand an Wörtern, aber auch an Mehrwortausdrücken, über den wir bereits verfügen, erheblich ausbauen; dies geschieht in Zusammenarbeit mit Göttingen einerseits, mit dem Leibniz-Institut für Deutsche Sprache andererseits. Darüber werden Sie in den kommenden Vorträgen Näheres hören.

Jetzt muss ich zum Schluss ein wenig aus der Textsorte fallen und zwei Leute loben. Eigentlich würde ich gerne sehr viele loben, denn sehr viele haben sich daran beteiligt. Aber das würde zu weit führen, und so greife ich zwei heraus. Der eine ist Martin Grötschel hier vorne, der – wenn dieses Oxymoron erlaubt ist – mit nüchterner Leidenschaft dieses Ganze betrieben hat, an dieses Vorhaben herangegangen ist. Das war ganz wunderbar, das können Sie sich nicht vorstellen. Ich rufe ihn irgendwann an und sage: „Kannst du nicht mal sehen, dass man das macht?“ Eine Stunde später ruft er zurück und sagt: „Also, ich habe den Sowieso nicht erreicht, aber ich habe mit der und der geredet. Machen wir.“ Oder manchmal auch: „Machen wir nicht.“ Das ist eine tolle Erfahrung, und ohne ihn wäre das einfach nicht gegangen. In diesem Raum gibt es noch eine weitere Person, ohne die es nicht gegangen wäre, das ist Alexander Geyken, der da drüben sitzt. Er ist seit vielen Jahren Arbeitsstellenleiter des *Digitalen Wörterbuchs der deutschen Sprache* und macht in dieser Funktion alles – wobei „macht alles“ auch heißt, dass er mit großem Geschick ganz hervorragende Mitarbeiter ausgesucht hat, die das Ganze miteinander tragen und die auch gemeinsam einen großen Teil der Arbeit des ZDL stemmen werden. Den beiden Genannten, aber auch allen anderen danke ich sehr herzlich. Es war bisher eine Freude, es wird auch in Zukunft eine Freude sein.

Wilhelm von Humboldt, dessen Elternhaus, wie Sie wissen, hier gestanden hat, Wilhelm von Humboldt hat in seiner ganz berühmten Denkschrift über die Höheren Wissenschaftlichen Anstalten gesagt: „Es ist daher eine Eigentümlichkeit dieser höheren Bildungsanstalten, dass sie die Wissenschaft immer als ein noch nicht ganz aufgelöstes Problem betrachten und daher immer im Forschen bleiben.“ Wenn diese Definition der wissenschaftlichen Tätigkeit zutrifft, dann ist unser Zentrum für digitale Lexikographie der deutschen Sprache eigentlich ein Exempel dafür. Es ist ein ganz großes Haus, das man unentwegt weiter ausbaut, das unser Wissen über die deutsche Sprache ständig dokumentiert und bereichert, das nicht zu einem Ende kommt, das aber jederzeit für jedermann, der sich für unsere Sprache interessiert, von Nutzen ist.



**Alexander Geyken** ist Arbeitsstellenleiter an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) für das ZDL und gemeinsam mit Wolfgang Klein wissenschaftlich verantwortlich für die lexikographische, technische und organisatorische Planung und Umsetzung des Arbeitsprogramms des Berliner Teilprojekts. Darüber hinaus ist er seit 2007 Arbeitsstellenleiter des Akademievorhabens „Digitales Wörterbuch der deutschen Sprache“ (DWDS) an der BBAW, einem Partnerprojekt des ZDL, sowie Mit-antragsteller in mehreren Drittmittelprojekten in den Bereichen Infrastruktur (CLARIN, CLARIAH, OCR-D) und Korpuslinguistik (t.evo). Als Sprachwissenschaftler hat er eine Privatdozentur an der Universität Potsdam inne.

## **DEN WORTGEBRAUCH VERLÄSSLICH UND UMFASSEND BESCHREIBEN: DIE ZIELE DER ZDL-PLATTFORM**

Sehr geehrte Damen und Herren, die Ziele des ZDL sind hochgesteckt. Dies wurde von Wolfgang Klein in dem vorangegangenen Vortrag motiviert und ausführlich beschrieben. Ich möchte die drei wichtigsten Ziele noch einmal kurz zusammenfassen. Erstens soll das ZDL den deutschen Wortschatz verlässlich und umfassend beschreiben, zweitens soll das ZDL leicht nutzbar und flexibel ergänzbar sein und schließlich sollen drittens die deutsche Hochsprache, aber auch ihre Varietäten, insbesondere regionale und stilistische Varianten, und die älteren Sprachstufen beschrieben werden.

In meinem Beitrag werde ich erläutern, wie diese Ziele während der projektfinanzierten Aufbauphase des ZDL umgesetzt werden sollen. In dieser zunächst für die Dauer von fünf Jahren und bei erfolgreicher Evaluierung für weitere drei Jahre geförderten Phase geht es um die Erarbeitung der Grundlagen des Zentrums: Diese umfassen zunächst die gesamte technische Infrastruktur und die Textkorpora, die ganz wichtig sind, wenn man die Verlässlichkeit bezüglich der

den Wortbeschreibungen zugrunde gelegten Quellen gewährleisten möchte. Im Zentrum der Arbeit steht jedoch die lexikographische Arbeit. Diese lässt sich in zwei Bereiche unterteilen, nämlich die Beschreibung der Gegenwartssprache und die Erarbeitung von Wortgeschichten. Für die Gegenwartssprache ist die BBAW als Koordinatorin des ZDL verantwortlich und kooperiert dabei im Teilbereich der Neologismen mit dem IDS. Dazu wird Herr Lobin im übernächsten Vortrag berichten. Und auch über den Bereich der Wortgeschichten werde ich in meinem Beitrag nichts sagen. Hier verweise ich auf den Vortrag von Herrn Harm, der im Anschluss an meinen Vortrag über die geplanten Aktivitäten der Göttinger Akademie berichten wird.

Die Frage, die uns in den letzten Wochen, als das ZDL sehr erfolgreich angekündigt wurde – es gab über 60 Presseveröffentlichungen –, von verschiedenen Journalisten immer gestellt worden ist, war, wie denn die umfassende lexikographische Beschreibung des deutschen Wortschatzes in der konkreten Projektarbeit aussehen soll. In der Tat ist die Erarbeitung eines Wörterbuchs von null an eine – gemessen an der Projektfinanzierung – zu zeitaufwändige Unternehmung. Wir erinnern uns an die Überschlagsrechnung von Herrn Klein, der für die Erarbeitung eines vollständigen Wörterbuchs mit der uns zur Verfügung stehenden Finanzierung weit mehr als 100 Jahre veranschlagt. Dies kann also nicht der Weg sein. Aus diesem Grunde stützen wir uns als Ausgangspunkt für den gegenwartssprachlichen Teil des ZDL auf die Quellen des *Digitalen Wörterbuchs der deutschen Sprache*, abgekürzt DWDS. Das DWDS besteht aus drei Säulen: Wörterbücher, Textkorpora sowie IT-gestützte Verfahren, mit Hilfe derer wir u. a. Wortstatistiken und dazugehörige Visualisierungen ableiten.

Das DWDS integriert mehrere bestehende gegenwartssprachliche Wörterbücher. Ausgangspunkt bildet das in der DDR in den 1960er und 70er Jahren erarbeitete *Wörterbuch der deutschen Gegenwartssprache* (WDG). Hinzu kommen Wortstrecken des 10-bändigen *Großwörterbuchs der deutschen Sprache* des Duden-Verlags in der Fassung von 1999 sowie kleinere Wortstrecken aus dem *Deutschen Wörterbuch* von Wahrig. Für diese Quellen verfügt die BBAW über die entsprechenden Rechte bzw. hat diese mit den Lizenzgebern vereinbart. Hinzu kommen das *Etymologische Wörterbuch des Deutschen* von Wolfgang Pfeifer, das *Deutsche Wörterbuch von Jacob Grimm und Wilhelm Grimm* sowie als Platzhalter für eine künftige wissenschaftliche Ontologie der *Open Thesaurus*. Perspektivisch wird auch die Integration von GermaNet angestrebt. GermaNet ist ein von der Universität Tübingen erstelltes lexikalisch-semantisches Netz

der deutschen Sprache, welches auf der Basis des weit verbreiteten Princeton WordNets erstellt wurde. Unsere eigene lexikographische Projektarbeit umfasst die Vereinheitlichung der Darstellung, die Überarbeitung bestehender Einträge sowie die Erstellung neuer Einträge. Dies ist eine fortlaufende Arbeit, der wir uns nun durch die Finanzierung des ZDL noch intensiver als bisher widmen können.

Die zweite Säule, die der Textkorpora, dient als empirische Grundlage für die Wörterbuchbeschreibung: Ein Gutteil der lexikographischen Arbeit besteht darin, die Bedeutungsbeschreibungen der Wortartikel aus den Verwendungen in den Textkorpora abzuleiten. Das DWDS verfügt über zwei Arten von Korpora: Referenzkorpora und Spezialkorpora. Bei den Referenzkorpora des DWDS handelt es sich um zeitlich und nach Textsorten ausgewogene Textsammlungen für den Zeitraum von etwa 1600 bis 2000. Für viele Zwecke der Wortschatzbeschreibung, insbesondere für etwas seltenere Wörter oder für den Fachwortschatz, sind diese Korpora zu klein. Wir haben diese daher durch Spezialkorpora ergänzt, die vor allem den Zeitraum der letzten 25 Jahre umfassen. Insgesamt verfügen wir derzeit über eine Sammlung von über 30 Millionen Dokumenten, im Wesentlichen Texte aus Tages- und Wochenzeitungen sowie aus Webkorpora. Dies entspricht in etwa einer Menge von 13 Milliarden Textwörtern, also 13 Milliarden aneinandergereihten Wörtern.

Auf der Startseite des DWDS, die unter [www.dwds.de](http://www.dwds.de) abrufbar ist, sind diese Säulen beschrieben und einzeln ansteuerbar. Ergänzt wird dies auf dem rechten Teil der Seite durch aktuelle Informationen aus dem Projekt, insbesondere auch durch das Wort des Tages. Dem Anlass gemäß haben wir heute das Wort „Auftakt“ gewählt. Es handelt sich dabei um ein relativ unscheinbares Wort, aber es ist in gewisser Weise typisch für unsere Arbeit, die darin besteht, den in unseren Wörterbuchquellen beschriebenen Wortschatz zu überarbeiten bzw. zu aktualisieren. Das heutige Wort des Tages, also das Wort „Auftakt“, bzw. sein Artikel, erfuhr für diesen Anlass eine Aktualisierung gegenüber seiner ursprünglichen Fassung, die dem *Wörterbuch der deutschen Gegenwartssprache* entnommen ist, wie man auf der Webseite der Quellsigle ablesen kann. Sehen wir uns den Artikel nun ein wenig genauer an. Die grammatischen Angaben wurden überprüft, die Vernetzung mit anderen Artikeln, beispielsweise den Komposita, die das Wort „Auftakt“ enthalten, vorgenommen. Diese kann man nun durch einen Klick in die Rubrik Wortbildung erreichen. Im Hauptteil des Wörterbuchs sind die verschiedenen Bedeutungen zusammen mit den Quellen beschrieben, aus denen die Bedeutungen entnommen sind. Belegt werden diese Bedeutungen durch Sätze aus unseren Korpora. In der rechten Spalte der Seite erfährt man

etwas über den Beginn der Verbreitung dieses Wortes im 19. Jh. bzw. über die „Karriere“ des Wortes in den letzten 50 Jahren. Darüber hinaus hat man die Möglichkeit, zu älteren Wörterbüchern zu springen bzw. sich die Korpus-Treffer anzusehen. Diese Nachschlagemöglichkeiten sollen perspektivisch für jedes Wort unserer Wörterbuchbasis zur Verfügung stehen.

Die Aufgaben und Herausforderungen des ZDL setzen nun genau darauf auf: Wir möchten Verlässlichkeit gewährleisten und gleichzeitig einen großen Artikelumfang bereitstellen. Dies gilt es für alle drei Säulen zu erreichen. Bezogen auf die jetzige Wörterbuchbasis, die sich, wie bereits erwähnt, aus verschiedenen Quellen großer Wörterbücher speist, gilt es zu beachten, dass dies nicht nur aktuell, sondern auch konsistent eingehalten werden muss. Die Herausforderung ist dabei die große Menge an Wörtern im Deutschen. Wolfgang Klein hatte schon berichtet: Sichtet man ein etwa eine Milliarde laufende Textwörter umfassendes Korpus, kommt man auf rund fünf Millionen verschiedener Wörter. Das steigt nicht ganz linear mit der wachsenden Korpus-Menge, aber aufgrund der reichhaltigen Möglichkeiten der deutschen Sprache, neue Wörter durch Zusammensetzungen zu bilden, lässt sich auch keine obere Grenze definieren. Die Beschreibung dieses „Wortwachstums“ wäre ein eigenes Thema. Bedeutend ist für uns aber an dieser Stelle, dass der deutsche Wortschatz die Stichwortanzahl auch der größten Wörterbücher der deutschen Gegenwartssprache weit übertrifft. Im Rahmen des ZDL bedeutet dies für uns, dass wir für den Umgang mit dieser Differenz zwischen der Gesamtheit des deutschen Wortschatzes und dem, was wir im Rahmen der Aufbauphase des ZDL lexikographisch beschreiben können, einen pragmatischen Weg der Umsetzung finden müssen. Dies heißt insbesondere, dass bei begrenzten Mitteln nicht alle Wörter in der gleichen lexikographischen Tiefe bearbeitet werden können.

Die zweite Herausforderung, der wir uns im Rahmen der Aufbauphase des ZDL stellen müssen, besteht in der systematischen Erweiterung der Textquellen. Denn wenn die Verlässlichkeit und damit auch die Akzeptanz bei allen Nutzerinnen und Nutzern des Systems gewährleistet sein soll, benötigen wir nicht nur eine große Anzahl an Texten, sondern auch nach Textsorten und Diskursen gut stratifizierte Textkorpora. An dieser Stelle möchte ich im Namen des ZDL an Verlage und Urheberrechtsinhaber appellieren, das ZDL mit Textspenden zu unterstützen, um damit die empirisch fundierte Beschreibung des deutschen Wortschatzes weiter zu verbessern.

Im Bereich der dritten Säule, der IT-gestützten Verfahren, wollen wir uns in den nächsten Jahren der verbesserten Automatisierung der Informationsextraktion aus Texten widmen. An dieser Stelle will ich nur ein Beispiel nennen, das aber besonders eindrücklich zeigt, wie sehr die Digitalisierung in die Lexikographie Einzug gehalten hat. Traditionell beruht die Auswahl geeigneter Satzzitate für Wortbedeutungen auf der sorgfältigen Lektüre der Texte. Als besonders gut gelten dabei Belege, die gewissen Kriterien genügen, die also beispielsweise besonders typisch und syntaktisch nicht zu komplex sind. Mit Methoden des maschinellen Lernens, die wir hierfür einsetzen wollen, ist es nunmehr möglich, dass die Systeme von den Bewertungen der Lexikographinnen und Lexikographen lernen, interaktiv, ohne dass hierbei explizit Kriterien angegeben werden müssen.

Zurück zur ersten Säule, der lexikographischen Arbeit. Hier möchte ich einige Zahlen nennen, die wir erreichen wollen. Das DWDS stellt bis zum Ende seiner Laufzeit 180.000 Einträge bereit und bringt diese in das ZDL ein. Weitere 20.000 Stichwörter zusätzlich muss das ZDL im Rahmen der Aufbauphase erstellen, um das im Antrag beschriebene Gesamtziel von insgesamt 200.000 Wörterbuchartikeln zu erreichen. Neben der Erstellung neuer, im Bestand noch nicht erfasster Wörterbuchartikel – dies können Neologismen sein ebenso wie Fachbegriffe – muss auch der Bestand in den Wörterbüchern auf Vollständigkeit und Richtigkeit hin überprüft werden. Dabei geht es nicht nur darum, Neubedeutungen nachzutragen, wie beispielsweise die Bedeutung von „Heuschrecke“ als „besonders renditeorientierter Investor“, sondern auch Bedeutungsveränderungen von vermeintlich unverdächtigen Wörtern, wie beispielsweise „Erosion“ in Zusammenhang mit Kollektiva oder Abstrakta, wie in „Erosion der Parteien“, „der Mittelschicht“ oder „des Rechtsbewusstseins“.

Bei etlichen Bereichen der Überarbeitung sind fundierte Fachkenntnisse notwendig. So trauen wir uns im Projektteam zwar zu, die Artikel „Hüpfburg“ oder „Auftakt“ zu beschreiben, aber von uns erstellte Definitionen von „Blockchain“ und „Amphetamin“ wollen wir vor der Veröffentlichung lieber von Fachleuten gegengelesen lassen. Hierzu haben wir im ZDL ein kleines Budget vorgesehen, welches die Beratung durch Fachleute vorsieht. Ein weiterer wichtiger Punkt der Überarbeitung betrifft die Frage, wie wir entscheiden, ob Artikel vollkommen neu zu erstellen sind oder ob eine partielle Überarbeitung ausreicht. An dieser Stelle machen wir uns die Vorteile der digitalen Arbeit zunutze. Ich möchte dies am Beispiel des bereits erwähnten Wortes „Auftakt“ noch einmal illustrieren. In diesem Falle haben wir entschieden, dass wesentliche Bestandteile des Artikels

des WDG aus dem Jahre 1967 heute nach wie vor gültig sind und somit in die Überarbeitung übernommen werden können, insbesondere die aus der Musik und aus der Verslehre stammende Bedeutung sowie eine daraus übertragene Bedeutung als „Einleitung für ein besonderes Ereignis“. Lediglich die daraus abgeleitete, speziell auf den Sport bezogene Bedeutung haben die Autorinnen und Autoren des WDG seinerzeit nicht aufgenommen. Dies haben wir nun nachgeholt und diese Bedeutung von „Auftakt“ im Sinne des „ersten in einer Reihe sportlicher Wettkämpfe“ nachgetragen sowie mit der Quellensigle ZDL 2019 zur Markierung der Urheberschaft versehen. Diese Art der minimalinvasiven Überarbeitung spart nicht nur Zeit, sie ermöglicht darüber hinaus den schnellen Überblick über sprachliche Veränderungen, da sich diese schnell an Quellensiglen ablesen lassen, die hinter jeder Bedeutungsdefinition angebracht sind. Bei dieser Gelegenheit möchte ich auch kurz auf die weiteren Informationen eingehen, die das DWDS für den Eintrag Auftakt auf der Webseite bereitstellt ([www.dwds.de/wb/Auftakt](http://www.dwds.de/wb/Auftakt)). Dies ist neben einem Thesaurus – derzeit setzen wir den kollaborativ erstellten *OpenThesaurus* ein –, über den man etwas über die Synonyme, aber auch die Ober- und Unterbegriffe erfährt, auch noch das sogenannte DWDS-Wortprofil. Dabei handelt es sich um eine mehrere Millionen Einträge umfassende Datenbank von typischen Wortverbindungen. Diese Datenbank kann auf der DWDS-Webseite durch Klick auf DWDS-Wortprofil nachgeschlagen werden. Sie wird entweder in Form einer allgemeinen Schlagwortwolke präsentiert oder – im Modus Fortgeschrittene – als Tabellenform, wobei jede Tabelle einer speziellen syntaktischen Kategorie entspricht. Für die beiden Kategorien Adjektiv und Präpositionalgruppe findet sich das in Abbildung 1.1 und 1.2.

hat Adjektivattribut	logDice ↓↑	Freq. ↓↑
4. verpatzt	8.4	80
5. fulminant	8.2	86
6. vielversprechend	7.7	81
7. verhalten	7.2	61
8. glanzvoll	7.2	54
9. mißglückt	7.1	49
10. enttäuschend	6.8	59
11. festlich	6.7	42
12. verregnet	6.6	24
13. mißlungen	6.6	23
14. durchwachsen	6.5	21

Abb. 1.1

hat Präpositionalgruppe	logDice ↓↑	Freq. ↓↑
1. nach Maß	8.3	263
2. in Oberstdorf	7.1	20
3. zu Feierlichkeit	6.8	16
4. zum Superwahljahr	6.5	10
5. zum Wahljahr	6.4	15
6. zum Jubiläumsjahr	6.4	11
7. zur Staffel	6.4	14
8. für Wahljahr	6.3	14
9. zu Veranstaltungsreihe	6.3	9
10. für Superwahljahr	6.3	9
11. zum Jubiläum	6.2	20

Abb. 1.2

Man findet hier, ich lese die ersten Einträge der Schlagwortwolke einmal vor, einen „furiosen“, einen „gelungenen“ oder einen „verheißungsvollen Auftakt“ unter der Rubrik Adjektiv-Nomen. Unter der Rubrik Präpositionalgruppe findet man besonders prominent den „Auftakt nach Maß“ – den wir heute hoffentlich auch haben werden. Wenn man übrigens auf eine Wortverbindung klickt, erhält man ausgewählte Korpusbelege, in denen diese Wortverbindungen vorkommen. Durch diese Rückbindung an authentische Satzkontexte können sich Nutzerinnen und Nutzer selbst vergewissern. Eine weitere Recherchemöglichkeit, die das System bietet, ist der sogenannte Wortvergleich. Dieser ist besonders dann von Interesse, wenn man Synonyme und deren Unterschiede untersuchen möchte. Als Beispiel möchte ich Ihnen hier „Auftakt“ im Vergleich zu „Beginn“ präsentieren. Diese beiden Wörter scheinen auf den ersten Blick sehr ähnlich. Durch unsere automatische Analyse lassen sich jedoch Verwendungsunterschiede sehr detailliert aufzeigen. In Abbildung 2 habe ich einmal nur die Adjektive genommen, mit denen die beiden Wörter typischerweise kombiniert werden. Die Adjektive in blauer Schrift sind vorwiegend typisch für „Auftakt“, diejenigen in roter Schrift für „Beginn“, die Zahlen, die unter der Spalte *logDice* stehen, geben die Stärke der Verbindung an. Je größer die Zahl, desto stärker die Wortverbindung. Der Graphik kann man daher entnehmen, dass die Adjektive „verregnet“, „misslungen“, aber auch „stimmungsvoll“ und „grandios“ besonders typisch für „Auftakt“ sind, wohingegen „zögernd“, „zaghaft“ und „frühestmöglich“ besonders typisch für „Beginn“ sind. Darüber hinaus, dies ist aber nicht in der Graphik abzulesen, gibt das System auch Adjektive aus, die in der Schnittmenge beider Wörter sind. So findet man beispielsweise in unseren Texten sowohl den „gelungenen Auftakt“ als auch den „gelungenen Beginn“. Die statistische Auswertung ergibt jedoch, dass „gelungen“ bedeutend häufiger mit „Auftakt“ verwendet wird.

Derlei Auswertungen lassen sich im DWDS-System auf Knopfdruck ermitteln. Fragestellungen nach typischen Wortverbindungen werden nicht nur im Bereich des Fremdsprachenlernens in der Produktionssituation nachgeschlagen. Wie man am Beispiel des Wortvergleichs sieht, lassen sich damit ebenfalls Dinge nachschlagen, die auch für Muttersprachler bei Formulierungsunsicherheiten sehr wertvoll sein können.

hat Adjektivattribut <sup>1</sup>	logDice	logDice	Freq.	Freq.
1. verregneten	6.6	-	24	-
2. mißlungenen	6.6	-	23	-
3. verkorksten	6.5	-	24	-
4. missratenen	6.5	-	20	-
5. stimmungsvollen	6.5	-	21	-
6. grandiosen	6.1	-	38	-
7. glänzenden	6.1	-	58	-
8. geglückten	6.0	-	18	-
9. zögerndem	-	6.0	-	15
10. zaghaften	-	6.0	-	19
11. frühestmöglichen	-	6.1	-	15

Abb. 2

Ich komme nun zu einem anderen Punkt, nämlich dem der Vokabularabdeckung. Dahinter verbirgt sich die folgende einfache Frage: Wie viele Wörter eines durchschnittlichen Zeitungsartikels sind nicht als Wörterbuchartikel im DWDS bzw. im *Duden* vermerkt? Erwärte würde man ja, dass jedes wichtige Wort in den Wörterbüchern beschrieben ist. Dem ist aber nicht so, wie ich an einem kleinen Beispiel illustrieren möchte. Hierzu habe ich einen Zeitungsartikel von vorgestern aus Focus Online ausgewählt mit dem Titel „Vor Gericht unbrauchbar: Überwachungskameras in deutschen Bahnhöfen liefern oft Schrott“ (*Focus Online* 27.1.2019). Diese Auswahl erfolgte nach keinem bestimmten Kriterium, ich hätte auch einen beliebigen anderen Artikel von diesem Tag auswählen können. Der von mir ausgewählte Artikel ist mit etwa 1.500 Wörtern länger als ein durchschnittlicher Online-Artikel. Aus Zeitgründen verzichte ich hier auf die komplette Auswertung, aber ich möchte kurz auf die folgenden fünf Wörter eingehen, die weder im DWDS noch im *Duden* verzeichnet sind. Es handelt sich dabei um die Wörter (1) „Überwachungsfilme“, (2) „Scheißtag“, (3) „Ersatzzug“, (4) „verdachtsunabhängig“ und (5) „Polizeideutsch“. Zwei dieser Wörter, nämlich (1) und (2), lassen sich von ihrer Bedeutung her gut erschließen, wenn man sie in ihre

Bestandteile zerlegt. Diese müsste man also nicht zwingend in ein Wörterbuch aufnehmen, wobei sich schon die Frage stellt, warum „Überwachungskamera“ im Wörterbuch beschrieben ist, „Überwachungsfilm“ aber nicht. (3) sollte hingegen in ein Wörterbuch aufgenommen werden. Dafür spricht nicht nur die verstärkte Verwendungshäufigkeit des Wortes in den letzten Jahren (vergleichen Sie hierzu die Wortverlaufskurve im DWDS), sondern auch, dass die Bedeutung usualisiert, also auf einen speziellen DB-Kontext begrenzt ist. Noch klarer ist der Fall bei „verdachtsunabhängig“ (4), welches oft in Zusammenhang mit Personenkontrollen oder (Daten-)Speicherung erwähnt wird. Es kommt zwar ursprünglich aus der Verwaltungssprache, hat aber mittlerweile ebenso in die Allgemeinsprache Eingang gehalten. Das Wort ist nicht selbsterklärend und sollte daher unbedingt auch in ein allgemeinsprachliches Wörterbuch aufgenommen werden. „Polizeideutsch“ (5) ist eine umgangssprachliche Variante für den Polizeijargon, sollte daher auch in das Wörterbuch aufgenommen werden.

Von diesen fünf fehlenden Wörtern würde man somit wenigstens (3), (4) und (5) neu in das Wörterbuch aufnehmen und beschreiben wollen. Hierfür haben wir im ZDL eine Datenbank angelegt, in die solche Neuaufnahmewünsche als Desiderata eingehen und dann nach einem vereinbarten Workflow abgearbeitet werden können. Derzeit umfasst diese Datenbank etwa 30.000 Einträge, davon etwa 40% fehlende Wörter, 25% fehlende oder veraltete Bedeutungen, der Rest verteilt sich auf andere, vornehmlich kleinere Fehlerklassen. Daran erkennt man, dass es sich bei den von mir identifizierten drei Wörtern um keine Einzelfälle handelt, sondern dass die Reichhaltigkeit der Sprache zu täglich neuen Desiderata führt. Die Datenbank wird übrigens derzeit von etwa 20 Personen des BBAW-Projektteams „befüllt“. Konkret erfolgt dies über ein Webformular auf der DWDS-Website.

Ein Bereich, auf den wir im ZDL aufgrund seines politischen Auftrags hinwirken werden, ist die Öffentlichkeitsarbeit. Hierfür wurde an der BBAW eine Stelle eingerichtet. Zunächst einmal bedeutete es für den Berliner Teil des ZDL, dass wir uns zeitnah mit aktuellen Wörterbuchartikeln an künftigen Wörtern des Jahres, dem Unwort des Jahres, dem Jugendwort des Jahres, auch dem Anglizismus des Jahres beteiligen wollen. Herr Stefanowitsch sitzt auch im Publikum, und ich glaube, etliche werden es mitbekommen haben, ich habe noch eigens gewartet: Das „Gendersternchen“ ist vor ein paar Stunden offiziell zum Anglizismus des Jahres gewählt worden. Diese Wahl wurde in vielen Medien heute publiziert. Und zeitgleich dazu gibt es das Wort mit einer lexikographischen Beschreibung jetzt auch im DWDS, es ist also seit heute nachschlagbar. Dies wurde dadurch ermöglicht,

dass das ZDL als Teil der Jury schon frühzeitig in die Wahl mit eingebunden wurde und daher den Artikel bereits vorbereiten konnte. Denn ehrlicherweise muss man sagen: Innerhalb von drei Stunden hätten wir es nicht geschafft, diesen Artikel online zu stellen. Perspektivisch wollen wir uns in der Öffentlichkeitsarbeit aber nicht nur auf die Erstellung von Wörterbuchartikeln im Zusammenhang mit prämierten Wörtern begrenzen. Wir planen auch die Einrichtung eines Blogs mit Miscellen zu aktuellen Wörtern, zu Etymologien sowie auch einfach zu Fragen der Rechtschreibung oder der Stilistik.

In den verbleibenden Minuten möchte ich Ihnen einige praktische Beispiele präsentieren, die zeigen, wie wir mit unserem System einzelne Wortkarrieren, aber auch Wortkarrieren im Vergleich nachzeichnen können. Ein besonders plakatives Beispiel ist der Vergleich von „Currywurst“, „Döner“ und „Bratwurst“. Als blaue Linie sehen Sie die Häufigkeitsvorkommen der „Currywurst“, über die bereits zwei Jahrzehnte vor dem „Döner“ in den Zeitungen geschrieben wurde. Der „Döner“ (rote Linie) kam erst Mitte der 1980er Jahre dazu und überholte seit etwa 2015 in der Auftretenshäufigkeit die „Currywurst“ – mit stark steigender Tendenz. Wenn man das mit der „Bratwurst“ (grüne Linie) vergleicht, die ja nach wie vor auf fast allen Jahrmärkten und Festveranstaltungen angeboten wird, so sieht man, dass sie noch leicht über dem „Döner“ steht. Aber man könnte bereits extrapolieren, dass das in ca. fünf Jahren nicht mehr so sein wird.

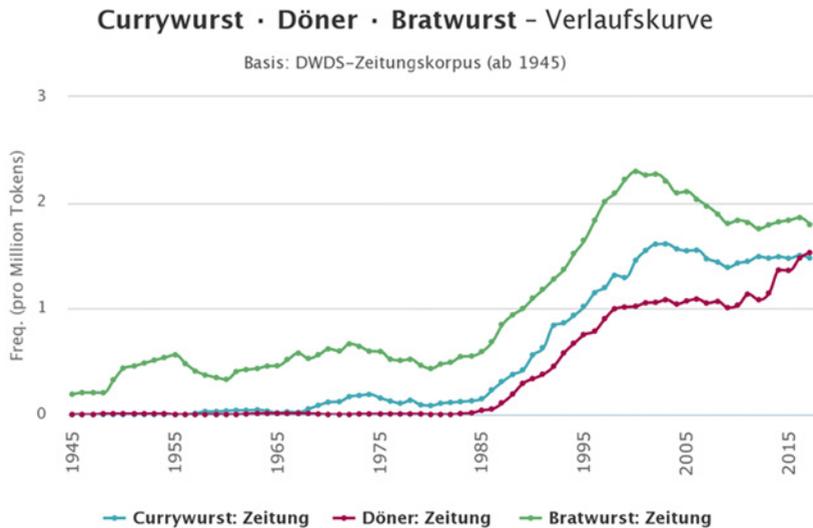


Abb. 3

Ähnliche Vergleiche habe ich mit „Turnschuh“ und „Sneakers“, „Fahrkarte“ oder „Ticket“ durchgeführt. Auch hier erkennt man, dass die jüngeren „Sneakers“ stark wachsen und bereits sehr nahe an die „Turnschuhe“ herangekommen sind. Bei „Fahrkarte“ oder „Ticket“ ist der Vergleich schon zugunsten des englischen Wortes entschieden. Es gibt zwar noch Fahrkartenautomaten bei der Bahn, aber genutzt wird schon seit geraumer Zeit nur noch das Wort „Ticket“. Perspektivisch wird damit die „Fahrkarte“ im alltäglichen Gebrauch allmählich veralten und vielleicht sogar ganz verschwinden. Bei anderen Wörtern wiederum ist man froh, dass sie nicht mehr so häufig gebraucht werden. Ein Beispiel hierfür ist die „Brotmarke“, unser Wort des Tages am 26. Januar 2019, welches an dessen starke Verwendung vor etwa 100 Jahren erinnert. Es kommt in unseren aktuellen Korpora kaum mehr vor.

Wir bekommen in diesem Zusammenhang auch immer wieder Nutzeranfragen, eine davon war: Ist „für gewöhnlich“ eigentlich gutes Deutsch? Wir konnten hier nicht nur nachweisen, dass diese Formulierung schon sehr lange in der deutschen Sprache verwendet wird, sondern auch, dass die den Ausdruck gebrauchenden Autoren durchaus namhaft sind. Auch wenn man „für gewöhnlich“ selber nicht nutzt, muss man diese empirische Evidenz anerkennen.

Den Möglichkeiten bei der Recherche von Wortkarrieren in unserem System sind kaum Grenzen gesetzt, als Nutzer der Seite können auch Sie alle in unseren Korpora verzeichneten Wörter oder Wortverbindungen in ihren Häufigkeitskurven vergleichen. Die Vergleiche sind hierbei fast immer noch interessanter als die Wortkarrieren einzelner Wörter selbst, weil sie sprachliche Konkurrenz abbilden und damit Indikatoren für Sprachwandel sein können.

Mein letztes Beispiel betrifft die „Auftaktveranstaltung“. Dies ist ein Wort, das wir im ZDL bislang lediglich ‚gebucht‘, aber noch nicht weiter lexikographisch aufgearbeitet haben. ‚Gebucht‘ bedeutet in der lexikographischen Werkstattsprache, dass ein Wort würdig für eine Aufnahme in das Wörterbuch ist, ohne dass es gleich weiter lexikographisch bearbeitet werden muss. Schlägt man dieses Wort im DWDS nach, erhält man bereits wichtige grammatische und statistische Informationen, aber noch keinen vollständigen Artikel. In der Wortverlaufskurve erkennt man einen „kamelartigen Höcker“, der für zwei zeitliche Intervalle steht, in denen das Wort Hochkonjunktur hatte. Dies war erstmals Mitte der 1980er Jahre und dann noch einmal um die Jahrtausendwende.

## Auftaktveranstaltung · Kickoff\* || Kick-off\* – Verlaufskurve

Basis: DWDS-Zeitungskorpus (ab 1945)

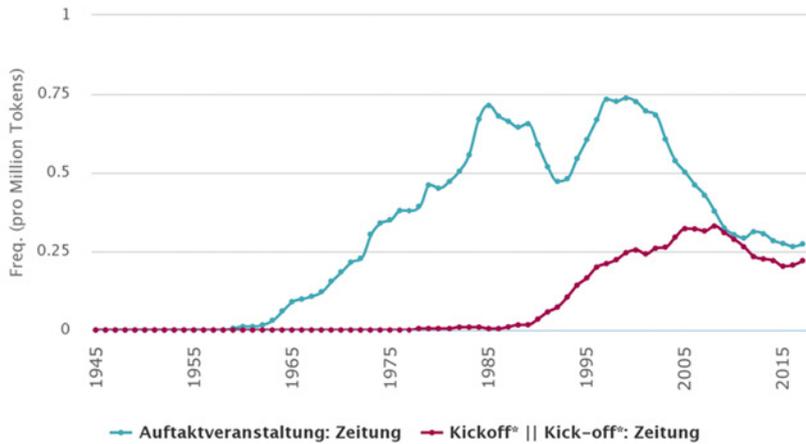


Abb. 4

Und auf meiner letzten Folie (s. Abbildung 4) habe ich einmal die „Auf-taktveranstaltung“ (in Blau) mit „Kick-off“ in den verschiedenen Varianten verglichen. Man erkennt das Anwachsen von „Kick-off“, aber noch hält die Auftakt-veranstaltung das jüngere „Kick-off“ knapp in Schach. Ich bitte Sie daher darum, in Berichten über diese Veranstaltung immer den Begriff „Auftaktveranstaltung“ zu verwenden – damit dies noch eine Weile so bleibt.

Vielen Dank für Ihre Aufmerksamkeit.



**Volker Harm** ist Lexikograph und Sprachwissenschaftler. Seine Forschungsschwerpunkte liegen im Bereich der Historischen Semantik und Grammatik sowie der Lexikologie. Er hat außerdem intensiv zur Schnittstelle Lexikon/Grammatik geforscht. Volker Harm lehrt als Privatdozent an der Georg-August-Universität Göttingen und ist Leiter der Göttinger Arbeitsstelle des ZDL, die mit dem Projekt „Wortgeschichte digital“ den neuhochdeutschen Wortschatz von ca. 1600 bis heute beschreibt. Er war als wissenschaftlicher Mitarbeiter und Arbeitsstellenleiter an der Neubearbeitung des *Deutschen Wörterbuchs* von Jacob und Wilhelm Grimm beteiligt.

## **„WORTGESCHICHTE DIGITAL“: AUFGABEN UND UMSETZUNGSSTRATEGIEN**

Woher ein Wort eigentlich kommt, was es früher einmal bedeutet und wie es sich zu dem entwickelt hat, was es heute ist – dies sind Fragen, die die Öffentlichkeit stark interessieren. Es handelt sich um Fragen wie beispielsweise: Woher kommt das schöne Wort „bärbeißig“, seit wann spricht man von „Immigration“ und wie hat man den Sachverhalt vorher benannt? Seit wann gebraucht man „Label“ neben „Etikett“? Und wie hängen eigentlich die Wörter „Etikett“ und „Etikette“ zusammen? Seit wann hat sich das Wort „Bildung“ auf Kosten von „Erziehung“ verbreitet? Für Fragen wie diese konsultieren Nutzer heute selbstverständlich das Internet. Eine Browser-Suche wie „Etymologie Etikett“ führt in aller Regel bereits auf die Informationsangebote des DWDS und damit auf eine verlässliche Auskunftsinanz.

Das über das DWDS verfügbare wortgeschichtliche Angebot steht, so könnte man sagen, im Wesentlichen auf drei Säulen: Die erste Säule bildet das von den Brüdern Grimm 1838 begonnene und 1960 abgeschlossene *Deutsche Wörterbuch*, dessen

32 Bände als Digitalisat im Internet durchsuchbar sind. (Die Neubearbeitung des *Deutschen Wörterbuchs*, die als Projekt 2016 abgeschlossen wurde, wird dieses digitale Angebot übrigens in absehbarer Zeit ergänzen.) Als zweite Säule kann man die historischen Textkorpora beschreiben, hier also die DWDS-Korpora, die das gesamte 20. Jh. und das 21. Jh. abdecken, sowie das *Deutsche Textarchiv* mit seinen rund viereinhalbtausend Werken vom 15. Jh. bis 1927. In diesen Sammlungen kann ein Nutzer eigenständig Textstellen für ein Suchwort recherchieren. Die dritte Säule neben dem *Deutschen Wörterbuch* und den Textkorpora ist das *Etymologische Wörterbuch* von Wolfgang Pfeifer, das eine erste Orientierung für die Herkunft des Grundwortschatzes bietet.

Auf den Seiten des DWDS wird somit ein reichhaltiges und hochwertiges Informationsangebot präsentiert. Mit dem Vorhaben „Wortgeschichte digital“, das an der Göttinger Akademie der Wissenschaften als einem der beiden Projektträger angesiedelt ist, wollen wir diesem Angebot gewissermaßen eine weitere Säule hinzufügen. Das Ziel dieses Projekts, das ich Ihnen im Folgenden etwas genauer vorstellen will, ist somit, wortgeschichtliche Informationen zu kulturell bedeutsamen Wörtern bereitzustellen, die bis jetzt noch gar nicht oder noch nicht hinreichend lexikographisch beschrieben worden sind. So sucht man vergeblich nach aktuellen historischen Informationen zu wichtigen Schlüsselwörtern wie den genannten Wörtern „Bildung“ oder „Erziehung“. Aber auch Wörter jüngeren Datums wie „Immigration“ sind bisher noch nicht oder noch nicht hinreichend historisch beschrieben. Das liegt zu einem großen Teil daran, dass es um die historische Wortforschung insgesamt nicht gut bestellt ist, dass Wortgeschichte trotz beachtlicher Forschungsergebnisse im Einzelnen innerhalb der stark auf die Grammatik ausgerichteten Sprachwissenschaften eher ein Schattendasein fristet. Es gibt auf diesem Gebiet somit enormen Nachholbedarf. Will die Sprachwissenschaft sich jedoch nicht vollends ins Esoterische verabschieden, will sie den Verlust öffentlichen Interesses für ihre Gegenstände nicht hinnehmen, so ist hier dringend Abhilfe zu schaffen – denn es gibt ohne Zweifel ein konstantes und breites Wissensbedürfnis im Hinblick auf wortgeschichtliche Fragen, jedenfalls ein breiteres als an vielen anderen Themen der Sprachwissenschaft. So verbindet sich denn auch mit dem Projekt „Wortgeschichte digital“ das Bestreben, ein zu lange schon brachliegendes wissenschaftliches Feld wieder zu bearbeiten und der Disziplin Wortgeschichte damit zu einer neuen Blüte zu verhelfen. Angesichts der digitalen Instrumente und Ressourcen, über die wir heute verfügen, kann man durchaus sagen, dass jetzt der richtige Zeitpunkt für ein solches Projekt gekommen ist, und mit dem ZDL hat es ganz sicher auch seinen richtigen Ort

in der Forschungslandschaft gefunden; denn mit den vier Akademien und dem Leibniz-Institut für Deutsche Sprache in Mannheim sind genau die Institutionen beteiligt, die über eine beachtliche Tradition der historischen Wortschatzforschung verfügen.

Im zunächst folgenden Abschnitt möchte ich versuchen, den Gegenstand des Vorhabens „Wortgeschichte digital“ kurz zu umreißen. In einem zweiten Teil will ich dann auf die Methoden eingehen, mit denen wir arbeiten wollen. Zunächst also zum Gegenstand des Projekts.

Wenn man den deutschen Wortschatz historisch beschreiben will, so hat man es im Grunde mit einer unlösbaren Aufgabe zu tun, und zwar zum einen, weil der Wortschatz des Deutschen wie der jeder anderen Kultursprache unermesslich groß ist, und zum anderen, weil sich Bestand und Bedeutung der Wörter ständig ändern. Wer eine umfassende Wortschatzbeschreibung vorlegt, muss also im Grunde, sobald er fertig ist, wieder von vorn anfangen – Wortschatzarbeit ist Sisyphos-Arbeit: Sobald der Felsbrocken auf den Berg gewälzt ist, rollt er auch schon wieder hinab. Wegen der Größe und der Flexibilität des Gegenstandes muss man bei jeder lexikographischen Beschreibung des deutschen Wortschatzes, sei sie gegenwartsbezogen oder historisch, auswählen, und zwar streng auswählen. Wir haben für unser Projekt eine Reihe von Auswahlkriterien festgelegt. Wir werden uns zunächst auf einen Zeitausschnitt beschränken, nämlich auf die jüngere Sprachgeschichte von ca. 1600 bis heute. Dies hat seinen Grund zunächst darin, dass die Epoche von 1600 bis in die Gegenwart, die man auch als das Neuhochdeutsche im engeren Sinne bezeichnen kann, wissenschaftlich nur sehr unzureichend bis gar nicht beschrieben ist. Für die Zeit vor 1600 kann man sich für wortgeschichtliche Fragen auf eine Reihe hervorragender Spezialwörterbücher zu den einzelnen Epochen stützen, also auf das *Althochdeutsche Wörterbuch*, das *Mittelhochdeutsche Wörterbuch* sowie das *Frühneuhochdeutsche Wörterbuch*. Für die Zeit von 1600 bis heute gibt es jedoch keine eigene lexikographische Darstellung. Dieser Zustand steht in deutlichem Kontrast zur Bedeutung dieser Zeit: Es handelt sich um die Epoche, in der unsere Gegenwart entscheidend geformt wurde und in der auch das Deutsche seine Prägung als voll funktionsfähige Kultursprache erfahren hat. Seit 1600 gibt es, wie jeder weiß, eine Explosion von Ideen, Erfindungen und Entdeckungen auf allen Gebieten des menschlichen Lebens, in diese Epoche fallen einschneidende historische Ereignisse und gesellschaftliche Umwälzungen; diese Zeit hat daher wie keine andere unseren heutigen

Wortschatz beeinflusst. Ohne eine vertiefte Kenntnis dieses Zeitraums können wir somit letztlich unsere Gegenwart nicht verstehen.

Neben der grundlegenden zeitlichen Einschränkung auf die Epoche von ca. 1600 bis heute gibt es mindestens einen weiteren Gesichtspunkt, der für unsere Wortschatzauswahl leitend ist. Wir haben uns entschieden, den Wortschatz nach Themenschwerpunkten zu bearbeiten, und wollen mit diesen thematischen Schwerpunkten gewissermaßen Schneisen in den Wortschatz schlagen. In den Sprachgeschichten des Deutschen können Sie nachlesen, dass die Zurüstung der deutschen Sprache für neue Themenfelder sowie die schrittweise sprachliche Aneignung von Sachgebieten, die vorher entweder gar nicht oder vorwiegend auf Latein oder Französisch verhandelt wurden, das zentrale Charakteristikum des Neuhochdeutschen als Epoche ist. Es sind also naheliegenderweise (aber keineswegs ausschließlich) die Wortschätze von Politik und Gesellschaft, Wissenschaft und Technik sowie der Kultur im engeren Sinne, die in dieser Zeit einen starken Wandel und einen umfassenden Ausbau erfahren haben. Themenfelder wie diese werden wir deshalb in den Mittelpunkt stellen. Wir werden uns also weniger mit der Gebäudebezeichnung „Haus“ befassen, zu der es wahrscheinlich nicht viel Neues herauszufinden gibt, was einen breiteren Nutzerkreis interessieren könnte, sondern wir werden uns eher um die semantischen Entwicklungen bei „Haushalt“ kümmern; ein Wort wie „Baum“ ist weniger wichtig für uns als „Baumsterben“, das Gegensatzpaar „öffentlich“ und „privat“ ist für uns wahrscheinlich von größerem Interesse als die Adjektive „warm“ und „kalt“, und Schlüsselwörter wie „Kultur“, „Bildung“, „Erziehung“, „Freizeit“ oder „Glück“ stehen (wie gesagt) ohnehin weit oben auf unserer Lemmaliste.

Soviel zum Gegenstand unseres Vorhabens. Mit dem Projekt „Wortgeschichte digital“ wollen wir aber nicht nur den Wortschatz ab 1600 beschreiben, wir wollen in der Beschreibung auch neue Wege gehen und innovative Umsetzungsstrategien für wordhistorische Aufgabenstellungen entwickeln. Zu einigen wichtigen Strategien – dies ist der angekündigte zweite, methodische Punkt – will ich im Folgenden etwas sagen.

Eine methodische Neuerung besteht darin, dass wir Wortgeschichte erzählen wollen. Damit folgen wir der Überzeugung, dass die Geschichte eines Wortes, wie jede Geschichte, erzählt werden muss, d.h. nach einer zusammenhängenden Darstellung in einem Fließtext verlangt. Wir weichen damit von der traditionellen Struktur von Wörterbucheinträgen ab, die ja meistens nur aus

nummerierten Bedeutungsangaben und einem mehr oder weniger großen Block von Verwendungsbeispielen bestehen. In der historischen Lexikographie, besonders auch im *Grimm'schen Wörterbuch*, ist die Vorstellung verbreitet, dass die Geschichte eines Wortes sich schon irgendwie selbst erzählt, wenn man einem Nutzer nur die Bedeutungsangaben und die Belegstellen an die Hand gibt. Eine bloße Auflistung des Inventars von historischen Verwendungen mit entsprechenden Verwendungsbeispielen ist aber kaum ausreichend – eine Liste von Wortbedeutungen ergibt eben noch nicht die Bedeutungsgeschichte eines Wortes, so wenig wie sich Geschichte aus einer chronologisch geordneten Aufzählung von Schlachten oder Wirtschaftskrisen ergibt.

Das Vorhaben, unsere Wortgeschichten in kleinen Texten darzustellen, folgt nun nicht allein der Überzeugung, dass Geschichte ihrem Wesen nach gar nicht anders dargestellt werden kann. Ein Vorteil des textuellen Verfahrens ist schlicht, dass es für eine bessere Verständlichkeit sorgt: Die Lexikographin/der Lexikograph muss das, was zu sagen ist, nicht erst in eine streng formalisierte lexikographische Struktur (die sog. ‚Mikrostruktur‘) verpacken und dann hoffen, dass die Nutzerin/der Nutzer diese Information dann wieder genauso entpackt, wie es gemeint war. Die von der Nutzerin/dem Nutzer verlangte Übertragungsleistung gewissermaßen aus dem ‚Lexikographischen‘ ins Deutsche entfällt, wenn Wortgeschichte direkt in einem Fließtext präsentiert wird.

Hier habe ich als Beispielartikel das bisher historisch noch nicht beschriebene Wort „Zuwanderung“ ausgewählt. In dem fortlaufend geschriebenen Text, aus dem ich hier einen kurzen Auszug wiedergebe, können viele Übergänge und Zusammenhänge sichtbar gemacht werden, die man in einer harten lexikographischen Mikrostruktur oftmals nur schwer ansprechen kann, in diesem Fall etwa das Changieren der Wortbedeutung zwischen verschiedenen Textsorten und Diskurskontexten.

„‚Zuwanderung‘ ist seit der ersten Hälfte des 19. Jahrhunderts in den Korpora belegt und ersetzt das ältere ‚Zuwandel‘ (1DWB 16, 897). Es bezeichnet die ‚Ansiedlung einer Bevölkerungsgruppe in einem Ort, Gebiet, Land‘ und wird zunächst häufig in Bezug auf Siedlungsbewegungen in der Vergangenheit angewandt: [...]. In der Folge etabliert es sich als Beschreibung für demografische Sachverhalte. Hier steht es zum Teil gleichbedeutend neben Einwanderung und scheint sich überwiegend auf Bevölkerungszuzüge vom Land in die Stadt zu beziehen. [...]“

Den erzählten Wortgeschichten, zumindest den längeren, soll auch eine Zusammenfassung beigegeben sein, die die Grundzüge der Gebrauchsentwicklung noch einmal auf den Punkt bringt.

„Das Wort ‚Zuwanderung‘ wird zuerst für Völkerbewegungen in der Antike gebraucht und findet sich Ende des 19. Jahrhunderts auch speziell mit Bezug auf die Landflucht. Seit Anfang des 20. Jahrhunderts erhält es Eingang in politische Kontroversen. In den seit den 1980er Jahren geführten einschlägigen Debatten spielt es vor allem als neutraler bzw. positiv besetzter Gegenbegriff zu dem umstrittenen Wort ‚Einwanderung‘ eine wichtige Rolle.“

Das Ansinnen, Wortgeschichten zu erzählen, mag recht traditionell klingen, und man kann sich zu Recht fragen, worin der digitale Aspekt des Vorhabens liegt – erzählte Wortgeschichten könnte man ja auch in einem Buch versammeln. Wir nehmen freilich das nachgestellte Attribut in der Projektbezeichnung „Wortgeschichte digital“ durchaus ernst. Digital ist das Projekt zunächst einmal in dem Sinne, dass wir mit digitalen Textkorpora arbeiten und ein computergestütztes Redaktionssystem verwenden, um die Artikel zu schreiben. Entscheidend ist aber etwas anderes: Die Wortgeschichten werden nicht isoliert präsentiert, sondern mit anderen Informationspositionen maximal verknüpft und für eine übergreifende Suche zugänglich gemacht – wie es eben nur in einem digitalen Informationssystem überhaupt möglich ist.

Die angesprochenen Verknüpfungen sind vielfältiger Art. Sie betreffen zum einen natürlich die Wortgeschichten untereinander: Wer eine Geschichte zu „Demokratie“ schreibt, wird unweigerlich in seinem Text auch Wörter wie „Wahl“, „Wahlrecht“, „Volk“ oder auch „Monarchie“ thematisieren, und da ist es zweckmäßig, auch per Link direkt auf die entsprechenden Wortartikel zu verweisen. Neben solchen internen Verlinkungen spielen aber auch externe Verweise eine wichtige Rolle. So wird man sich ohne Zweifel häufig auf eine Beschreibungsposition im *Grimm'schen Wörterbuch* oder in einem der genannten Epochenwörterbücher per Link beziehen wollen. Eine wichtige Form von Verweisen betrifft natürlich auch das Korpus. In vielen herkömmlichen Wörterbüchern ist es so, dass Belegzitate in eine lexikographische Beschreibung eingeschoben sind; nicht selten stehen mehrere Belegzitate mitten in einem Satz, der dann erst nach einem mehrere Zeilen umfassenden Belegblock weitergeführt wird. In einer Wortgeschichte, die als digitaler Text präsentiert wird, kann dies leicht entzerrt werden, indem man in den Text Verweise in ein dem Text quasi hinterlegtes Belegkorpus einfügt.

Durch solche Korpusverweise wird der Lesefluss nicht gestört, der Nachweis kann z. B. per Klick oder Mouseover-Fenster nachgeschlagen werden. Des Weiteren wäre als spezifische Verknüpfungsoption auch die Möglichkeit zu nennen, einschlägige Daten soweit wie möglich mit vorhandenen Normdatensätzen zu verbinden.

Neben den internen und externen Verknüpfungen als spezifisch digitale Möglichkeiten möchte ich hier auch Visualisierungen nennen. Neben Bildern und Karten denke ich vor allem an Modellierungen von Häufigkeitsverteilungen und Häufigkeitsentwicklungen im Textkorpus; dies steht ja im Instrumentenkasten des DWDS bereits jetzt zur Verfügung (zum Beispielwort „Zuwanderung“ vgl. dazu die Wortverlaufskurven in Abb. 1 und 2). Zu denken wäre aber auch an Zeitleisten, welche die Entfaltung eines historischen Bedeutungsspektrums sichtbar machen können. Alles in allem präsentiert unser Projekt damit erzählte Wortgeschichten als Kernbestandteile eines vielfältigen und hochintegrierten Informationsangebots, wie es eben nur im Internet möglich ist.

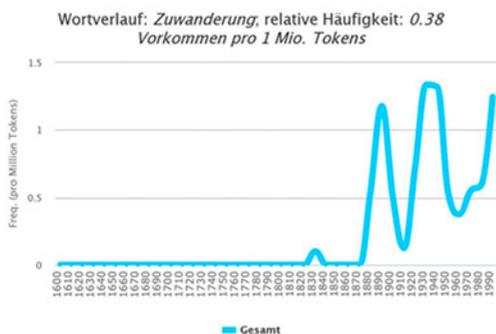


Abb. 1: Wortverlaufskurve „Zuwanderung“ (Stand 26.1.2019)

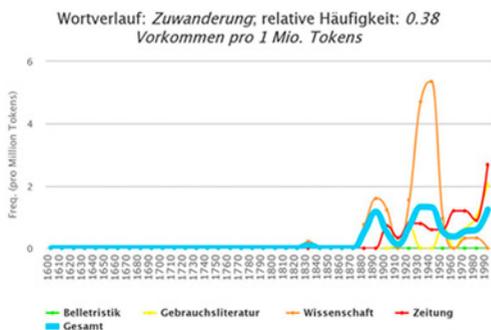


Abb. 2: Wortverlaufskure „Zuwanderung“ nach Textgattungen (Stand 26.1.2019)

In der Aufbauphase des ZDL ergeben sich somit zusammengefasst die folgenden Aufgaben für das Göttinger Teilprojekt: Wir wollen in erster Linie das Format der erzählten Wortgeschichte weiterentwickeln und gewissermaßen zur Reife bringen. Dazu gehört, dass Wortgeschichten systematisch abfragbar gemacht sowie inhaltlich und technisch mit den Visualisierungen und den anderen Elementen des Systems verbunden werden. Die Ergebnisse sollen dann in die ZDL-Plattform integriert werden. Die wesentlichen Schritte, die wir dabei gehen, sollen mit der Fachcommunity abgestimmt werden. Eine grundlegende Aufgabe wird dabei natürlich sein, an einer hinreichend großen Zahl von Artikeln zu demonstrieren, dass das Konzept funktioniert.

**Henning Lobin** ist Direktor des Leibniz-Instituts für Deutsche Sprache und Professor für Germanistische Linguistik an der Universität Mannheim. Zuvor war er Professor für Angewandte Sprachwissenschaft und Computerlinguistik an der Justus-Liebig-Universität Gießen und Direktor des dortigen Zentrums für Medien und Interaktivität. Im Rahmen seiner Forschung befasst er sich unter anderem mit den Auswirkungen der Digitalisierung auf die Sprache sowie korpuslinguistischen wie texttechnologischen Fragestellungen.



Foto: T. W. Klein

## **NEOLOGISMEN IM KONTEXT DER LEXIKALISCHEN FORSCHUNG AM LEIBNIZ-INSTITUT FÜR DEUTSCHE SPRACHE**

Meine sehr geehrten Damen und Herren, lieber Herr Grötschel, lieber Herr Gardt, lieber Herr Klein, zunächst einmal möchte ich Ihnen im Namen des Leibniz-Instituts für Deutsche Sprache (IDS) ganz herzlich dazu gratulieren, dass das Zentrum für digitale Lexikographie der deutschen Sprache nun endlich seine Arbeit aufnehmen kann, und Ihnen dafür alles Gute wünschen. Ich freue mich, dass das IDS dabei mit Ihnen kooperiert. IDS und BBAW sind ja schon seit geraumer Zeit über verschiedene Projektzusammenhänge miteinander verbunden, so etwa im Verbundvorhaben CLARIN-D, das seit Jahren vom Bundesforschungsministerium für den Aufbau einer sprachwissenschaftlichen Forschungsinfrastruktur gefördert wird. Und ich bin auch froh, dass Staatssekretär Rachel vorhin den Zusammenhang mit der Nationalen Forschungsdateninitiative hergestellt hat, denn dabei handelt es sich in der Tat um eine wichtige Thematik, mit der wir uns gemeinsam mit Erhard Hinrichs, der CLARIN-D als Sprecher leitet, gegenwärtig intensiv befassen, um die Entwicklung einer datenorientierten Sprachwissenschaft durch eine Forschungsdateninfrastruktur langfristig abzusichern.

Das IDS ist eine Einrichtung, die sich mit der Dokumentation und Erforschung der deutschen Sprache in Gegenwart und jüngerer Vergangenheit befasst und dabei Grundlagenforschung in verschiedenen systemlinguistischen Bereichen durchführt. Neben den etablierten Abteilungen Grammatik, Lexik und Pragmatik haben wir zu Beginn des Jahres 2019 eine neue Abteilung „Digitale Sprachwissenschaft“ geschaffen. Seit der Gründung des IDS in den 1960er Jahren gab es überdies die Ausrichtung auf Forschungsdaten, d. h. auf Korpora und andere Arten von Sprachdaten. Neben dem Bereich der Neologismen sehe ich hier einen wichtigen Schwerpunkt in der zukünftigen Kooperation mit dem ZDL, da wir am IDS mit dem Deutschen Referenzkorpus DeReKo einen Korpusbestand besitzen, der mittlerweile ca. 45 Milliarden Wörter umfasst.

Heute möchte ich jedoch die Neologismenforschung in den Kontext der lexikalischen, sprich: lexikographischen und lexikologischen Arbeit am IDS einordnen. Dieser Forschungsbereich ist verortet in der Abteilung Lexik, der Prof. Stefan Engelberg als Abteilungsleiter vorsteht, und gegliedert in drei Programmbereiche: „Lexikographie und Sprachdokumentation“ (Dr. Annette Klosa-Kückelhaus), „Lexikalische Syntagmatik“ (Prof. Stefan Engelberg) und „Lexik empirisch und digital“ (Prof. Carolin Müller-Spitzer).

Was die verschiedenen Projekte in diesen Programmbereichen verbindet, ist die Fokussierung der Dynamik des Lexikons. Das Lexikon wird dabei nicht als eine statische Gegebenheit angesehen, sondern es wird im Sinne einer in der Zeit verlaufenden Wortschatzentwicklung eine ständige Veränderung des Lexikons angenommen und nachgezeichnet, welche Eigenschaften derartiger Entwicklungsdynamiken aufweisen. Dieses Kernthema führt zu einer ganzen Reihe von grundlegenden Forschungsfragen, zum Beispiel die lexikologische Analyse und die lexikographische Dokumentation unzureichend erforschter Wortschatzbereiche; Neologismen sind einer davon. Weiterhin ist die Entwicklung von Repräsentationsformaten, Visualisierungs- und Explorationswerkzeugen zu nennen, die den Methoden und den Möglichkeiten insbesondere einer digitalen Vermittlung gerecht werden. Drittens ist die Identifikation und Beschreibung von syntagmatischen Mustern im Lexikon wichtig, also von über das einzelne Wort hinausgehenden Strukturen, die ihre ganz eigenen Charakteristika aufweisen. Und schließlich sollten auch empirische Studien zu der Frage durchgeführt werden, wie eigentlich mit Wörterbüchern, insbesondere mit digitalen Wörterbüchern, konkret umgegangen wird und welche Herausforderungen sich hier stellen – also die Benutzerperspektive.

Ich beginne damit, dass wir uns einige Beispiele für die Erschließung unzureichender Wortbestände ansehen. Neben Neologismen müssen wir dabei auch Paronyme berücksichtigen, also Wortpaare, die einander sehr ähnlich, aber nicht gleichbedeutend sind, „effektiv“ und „effizient“, „Technologie“ und „Technik“ etwa. Weiterhin spielt die Lexik des gesprochenen Deutsch eine wichtige Rolle, ein ausgesprochen interessantes Gebiet, da wir in diesem Bereich ganz andere lexikalische Zusammenhänge zu dokumentieren haben, als es in der Schriftsprache der Fall ist. Dabei liegt eine besondere methodische Herausforderung darin, dass wir hier bei weitem nicht im gleichen Maße durchsuchbare Korpora zur Verfügung haben, da zur Untersuchung der gesprochenen Sprache audiovisuelle Aufzeichnungen zunächst ins Schriftliche zu transkribieren sind, was eine überaus zeitintensive Tätigkeit darstellt. Schließlich die Fremdwörter: Wir untersuchen einerseits Fremdwörter, die ins Deutsche eingewandert sind, in einem Langfristprojekt zur Erarbeitung einer Neuauflage des *Deutschen Fremdwörterbuchs* etwa. Aber es geht uns auch um Fremdwörter, die vom Deutschen in eine andere Sprache übernommen worden sind. Diese Forschungsthemen werden korpusbasiert bearbeitet und sind von Benutzungsforschung flankiert.

Ein für das ZDL besonders relevantes Beispiel für die am IDS praktizierte Herangehensweise bilden die bereits genannten Neologismen. Wir verfolgen die Entwicklung von Neologismen in der deutschen Sprache kontinuierlich seit den 1990er Jahren und stellen unserer Untersuchungsergebnisse seit geraumer Zeit auch als digitales Wörterbuch ins Netz, jeweils in Tranchen von denjenigen Wörtern, die in einem bestimmten Zeitraum lexikographisch neu beschrieben wurden (s. <https://www.owid.de/docs/neo/start.jsp>). Das geschieht selbstverständlich nicht auf der Grundlage nur eines einmaligen Vorkommens eines neuen Wortes, sondern auf der Grundlage eines längeren Untersuchungsverlaufs, bei dem eine gewisse Stabilität zu verzeichnen ist, auch wenn diese sich in einer nächsten Phase durchaus auch wieder abschwächen kann. Wenn man sich die Listen von Neologismen in den verschiedenen Zeitscheiben – also aus den 90er Jahren, aus den Nullerjahren, aus den 10er Jahren – genauer ansieht, dann kann man eine Art Fingerabdruck der jeweiligen Epoche erkennen. Ein Begriff wie „Duckface“, der zur neuesten Tranche des *Neologismen-Wörterbuchs* gehört und eine Art Schnute bezeichnet, die man bei einem Selfie zieht, kommt in einer bestimmten Frequenz tatsächlich in unseren Korpora vor, also in den Zeitungs-, Zeitschriften- und sonstigen Texten, die darin enthalten sind.

Die lexikographisch erfassten Neologismen werden von uns auch mit Bedeutungsangaben versehen. Weitere Informationen lassen sich im Neologismenwörterbuch des IDS ebenfalls finden. Dieses Wörterbuch ist in einer Weise aufgebaut, dass man sich die Wörter nicht nur nach Jahren oder Jahrzehnten alphabetisch sortiert anschauen kann, sondern auch in alternativen Anordnungen. Wir sind insbesondere daran interessiert, die Dynamik der Neologismen sichtbar zu machen. Wenn Sie sich sehr bekannte Neologismen ansehen wie die Wörter „Mobiltelefon“ und „Handy“, dann lässt sich erkennen, dass 2015 in unserem sehr großen Deutschen Referenzkorpus 237.000 Vorkommen zu verzeichnen sind, aber auch eine kleine Anzahl des Wortes „Handtelefon“. Interessant ist es nun, wie sich diese Vorkommenshäufigkeiten über einzelne Jahresschnitte hinweg verteilen. Dabei ist eine Phase der Konsolidierung von „Handy“ gegenüber „Mobiltelefon“ zu erkennen. Das Wort „Mobiltelefon“ verschwindet allerdings keineswegs völlig, während der Alternativbegriff „Handtelefon“ Anfang bis Mitte der 1990er Jahre eine kurze Konjunktur erlebte, bis er dann sehr schnell wieder gänzlich ungebrauchlich wurde.

Interessant ist es auch, sich ganze linguistische Teilbereiche anzusehen, beispielsweise Wörter einer bestimmten Wortart, die als Neologismen in die deutsche Sprache kommen. In unserem Neologismen-Portal haben wir dafür Zugangsmethoden geschaffen, mit Hilfe derer man diese explorieren kann, beispielsweise neue Verben, die aus dem Englischen kommen: „adden“, „bladen“, „bloggen“, „bouldern“, „buzzen“, „carven“, „casten“, „chatten“, „chillen“ oder „daten“. Um eine derartige Abfrage vornehmen zu können, muss man bei den einzelnen Wörtern eine ganze Reihe von linguistischen Metadaten systematisch erfassen. Auch Inhaltsbereiche, die bei den jeweiligen Wörtern erfasst werden, können dazu herangezogen werden, etwa aus den Bereichen Ernährung, Bankwesen, Politik, Gesellschaft, Arbeitswelt usw. Wenn man das noch mit einer zeitlichen Kennzeichnung in Verbindung setzt, kann man beispielsweise erkennen, wann und in welcher Weise die Bereiche Ernährung oder Telekommunikation eine Schwerpunktbildung bei den Neologismen aufweisen. Man könnte sagen, dass dieses dynamische Lexikon uns geradezu eine Art Fingerabdruck einer Zeit präsentiert und dadurch auf der Grundlage weitergehender Auswertungsverfahren interessante Anknüpfungspunkte für die weitere Erschließung des Wortbestandes geben sind.

Einen zweiten wichtigen thematischen Komplex im Bereich lexikalischer Forschung am IDS bildet die Entwicklung von neuen Repräsentationsformaten,

Visualisierungen und Explorationswerkzeugen. Das Lehnwortportal beispielsweise ist ein Portal, das nicht etwa die Lehnwörter in der deutschen Sprache aufzeigt, sondern deutsche Wörter, die als Lehnwörter in andere Sprachen übernommen wurden (s. <http://lwp.ids-mannheim.de/>). Dort werden ganze Ketten von Übernahmeprozessen über mehrere Sprachen hinweg nachgezeichnet. Auf der IDS-Plattform OWID, dem *Online-Wortschatz-Informationssystem Deutsch*, werden verschiedene Spezialwörterbücher über eine einheitliche Oberfläche für die übergreifende Recherche angeboten (s. <https://www.owid.de/>). Besonders forschungsnahe oder innovative Visualisierungs- und Auswertungsmethoden, die oft auch weitergehende interaktive Komponenten aufweisen, werden über das Portal OWIDplus zur Verfügung gestellt (s. <https://www.owid.de/plus/>).

In einem dritten Bereich schließlich befassen wir uns mit sogenannten syntagmatischen Mustern im Wörterbuch und mit ihrer Identifikation und Beschreibung. Dies kann etwa die Argumentstruktur sein, die ein Verb um sich herum aufbaut, oder redereinleitende Ausdrücke – ein Thema, das am IDS gegenwärtig im Rahmen eines DFG-Projekts untersucht wird.

Einen weiteren Bereich bilden usuelle Wortverbindungen, genauer: lexikalische Muster, die eine gewisse Produktivität besitzen. Derartige Muster lassen sich variabel nutzen, und diese variable Nutzbarkeit wiederum ist von bestimmten Parametern abhängig. Diese Parameter herauszufinden ist eine ausgesprochen schwierige Aufgabe, denn es ist keineswegs klar, woran es liegt, dass manche Muster an einer bestimmten Stelle sehr produktiv sein können und an einer anderen Stelle nicht.

Ein Beispiel dazu: Baden-Württemberg hat sich vor zwanzig Jahren dazu entschieden, einen schnell sehr bekannt gewordenen Slogan zur Werbung für das Bundesland zu verwenden: „Wir können alles. Außer Hochdeutsch“. Selbstverständlich wird damit auf das Schwäbische angespielt. Interessant ist es nun, sich anzuschauen, wie dieses Beispiel abgeändert wird: „Ich kann alles. Außer Schwäbisch“, „Wir können alles. Außer Bahnhof“ (mit Bezug auf Stuttgart), „Wir können alles. Außer Fußball“ usw. Diese Varianten gibt es zu dem ursprünglichen Slogan in großer Zahl. Das Spektrum wird noch größer, wenn man nach dem zugrundeliegenden Muster auch in Korpora sucht und syntaktische Varianz berücksichtigt. Dabei findet man beispielsweise Relativsätze, in denen auf das Muster Bezug genommen wird, ohne dass dieses dabei als Landesslogan explizit sichtbar wird. Über das Wort „bekanntlich“ wird hier allerdings eine Art Redewiedergabe

signalisiert: „Und so würden Württemberger, die bekanntlich alles können außer Hochdeutsch, Berlin eines wünschen: die ganz große Kehrwoche ganzjährig.“ Eine weitere Variante aus der Tageszeitung *Mannheimer Morgen*: „Wir können alles außer Mainstream.“ Und das letzte Beispiel stellt eine Variante dar, bei der es fast schon schwerfällt, das indirekte Zitat zu erkennen, wenn man den Zusammenhang nicht kennt: „Günther Öttinger kann alles außer Englisch, was aber kein großer Schaden ist.“

Wir haben in unseren Korpora 256 Vorkommen des originalen Slogans gefunden und 251 Varianten davon, also eine fast gleich große Anzahl. Wenn man sich ansieht, wie sich diese Zahlen im zeitlichen Verlauf entwickeln, dann stellt man fest, dass es gar nicht lange dauert, bis die ersten Varianten erscheinen – die Variation setzt vielmehr sofort ein. In manchen Jahren übersteigen sogar die Varianten den Originalslogan in ihrer Häufigkeit. Es handelt sich um ein ganz typisches Phänomen, das sich mit derartigen Musterverwendungen zeigt, und dieses Phänomen weist eine ganz bestimmte Systematik auf.

Auch auf der Wortebene lassen sich vergleichbare Phänomene verzeichnen. Die Produktivität von Farbwörtern etwa ist in unseren Korpora stark ausgeprägt, für „schwarz“ etwa in folgender Weise: „achatschwarz“, „brokatschwarz“, „dohlen-schwarz“, „friedhofsschwarz“, „gelbblauschwarz“, „makaberschwarz“ usw. Bei „weiß“ findet man Wörter wie „apfelblütenweiß“, „diamantweiß“, „eisblauweiß“, „gänsefederweiß“ oder „joghurtweiß“. Interessant ist es nun, sich anzuschauen, wie die Produktivitätsrate bei diesen beiden Farbwörtern aussieht. Dabei kann man unter Nutzung eines Produktivitätsmaßes wie etwa dem von Harald Baajen (bei dem die Hapax legomena zu den Gesamtvorkommen in Beziehung gesetzt werden) feststellen, dass die „schwarz“-Produktivität erheblich, nämlich mehr als dreimal größer ist als die Produktivität von „weiß“.

Unseren vierten lexikalischen Forschungsbereich stellen die Nutzungsstudien dar, also empirische Studien zum Umgang mit Wörterbüchern und lexikalischen Online-Ressourcen, was schon seit langer Zeit am IDS betrieben wird. Wir besitzen sehr viel Erfahrung in der Anwendung unterschiedlicher Methoden, ob es sich um Fragebögen, Logfile-Studien, Eye Tracking- oder Screen Recording-Verfahren handelt. Was sich dabei für Herausforderungen stellen, soll auch hier an einem abschließenden Beispiel dargestellt werden.

Bei der Untersuchung der Zugriffszahlen auf den Eintrag „larmoyant“ im Online-Wörterbuch *Wiktionary* fiel uns auf, dass zu einem ganz bestimmten Zeitpunkt, nämlich am 6. Februar 2013 zwischen 22:00 und 23:00 Uhr, mehr als 1.500 Zugriffe auf das Wort zu verzeichnen waren, während in der Stunde zuvor nur fünf Besuche und in der Stunde danach lediglich 32 Besuche registriert wurden. Woran hat das gelegen? Die Mitarbeiterinnen und Mitarbeiter des IDS konnten sich dieses Phänomen lange nicht erklären. Es wurde in alle möglichen Richtungen recherchiert, bis schließlich die folgende Erklärung gefunden wurde: Am 6. Februar 2013 fand ein Freundschaftsspiel der Fußball-Nationalmannschaften von Deutschland und Frankreich statt. Dabei wurde genau um 22:23 Uhr vom Kommentator Folgendes geäußert: „Der ist jetzt aber richtig sauer. Das passt dem gar nicht. Und wenn ich es richtig sehe, blickt er da hin zu Toni Kroos. Das ist ihm ein bisschen zu larmoyant.“ Mit solchen Effekten wird man also konfrontiert, wenn man sich die realen Nutzungsdaten von Onlinewörterbüchern ansieht und versucht, sich einen Reim darauf zu machen.

Ich danke Ihnen allen für Ihre Aufmerksamkeit und danke auch meinen Kolleginnen und Kollegen Stefan Engelberg, Annette Klosa-Kückelhaus und Carolin Müller-Spitzer für die Unterstützung bei der inhaltlichen Erstellung dieses Vortrags.



Alle Fotos: BBAW, Judith Affolter



**Alexandra N. Lenz** ist Professorin für Germanistische Sprachwissenschaft (Sprachgeschichte und Varietätenlinguistik) an der Universität Wien. Zudem ist sie Wirkliches Mitglied der Österreichischen Akademie der Wissenschaften und fungiert dort als stellvertretende Direktorin des Austrian Centre for Digital Humanities (ACDH) sowie als Leiterin der Forschungsabteilung „Variation und Wandel des Deutschen in Österreich“. Ihre Forschungsschwerpunkte sind Variationslinguistik, Spracheinstellungsforschung, Syntax, kognitive Semantik sowie Lexikologie und Lexikographie.



Foto: BBAW, Judith Affolter

## VARIATIO DELECTAT?

### ZUM VERHÄLTNIS VON LEXIKOGRAPHIE UND SPRACHLICHER VIELFALT

#### 1. Einleitung

Der folgende Beitrag ist im Spannungsfeld von Lexikographie und Variationslinguistik angesiedelt. Im Zentrum steht das komplexe Verhältnis von sprachlicher Vielfalt im Sprachgebrauch einerseits und Normierung bzw. Deskription des/eines Sprachsystems in lexikographischen Nachschlagewerken andererseits. Dem Anspruch des ZDL entsprechend wird das Diskussionsobjekt eingeschränkt auf (v. a. standardsprachlich) ausgerichtete lexikographische Nachschlagewerke, wie das ZDL aktuell eines in Form einer digitalen Informationsplattform plant, und welche – so das ambitionierte Ziel –, „den deutschen Wortschatz umfassend und verlässlich beschreibt und das diese Beschreibung nutzergerecht ausgestaltet“, denn: „Verlässliche Informationen über Schreibweise, Aussprache, Bedeutungen, grammatische Eigenschaften, regionale wie stilistische Variationen und Herkunft der Wörter des Deutschen sind für WissenschaftlerInnen vieler Fächer, aber auch

für JournalistInnen, ÜbersetzerInnen, LehrerInnen, DeutschlernerInnen aus aller Welt, überhaupt für jeden, der sich für die deutsche Sprache interessiert, von Bedeutung.“ (<https://www.zentrum-lexikographie.de/ueber-das-zdl.html#traeger>) [letzter Zugriff: 19.08.2019]). Diese Zielsetzung wirft verschiedene Fragen auf, die im Folgenden einer ausführlichen Diskussion unterzogen werden sollen. (Nebenbemerkung: Aspekte der forschungspraktischen Machbarkeit umfangreicher Variationsdokumentation und überhaupt auch der dieser vorgelagerten Erhebung und Analyse von Variationsphänomenen, die in diesem Fragekomplex natürlich eine sehr entscheidende Rolle spielen, seien hier nur erwähnt, aber nicht weiter ausgeführt.)

1. *Wie viel Variation braucht ein Wörterbuch?* Das heißt, wie viel Information bzgl. der im Sprachgebrauch der Sprachgemeinschaft(en) einer Gesamtsprache vorherrschenden Variation muss in einem Wörterbuch lexikographisch aufbereitet und dokumentiert werden, und dies insbesondere in einem Wörterbuch, das den Anspruch erhebt, eine verlässliche Informationsquelle zu sein im Hinblick auf „Schreibweise, Aussprache, Bedeutungen, grammatische Eigenschaften, regionale wie stilistische Variationen und Herkunft der Wörter“ (<https://www.zentrum-lexikographie.de/ueber-das-zdl.html#traeger>) [letzter Zugriff: 19.08.2019])?
2. *Wie viel Variation verträgt ein Wörterbuch?* Das heißt, welches quantitative Ausmaß und welcher Komplexitätsgrad der Variationsdokumentation ist für ein Nachschlagewerk zuträglich bzw. für die NutzerInnen tragbar, um den angestrebten (und möglicherweise nicht selten auch konkurrierenden) Zielen der Vollständigkeit, Verlässlichkeit und Nutzerfreundlichkeit gerecht zu werden?
3. *Wie stellt ein Wörterbuch Variation „optimal“ dar?* Das heißt: Gesetzt den Fall, LexikographInnen entscheiden sich dafür, Sprachvariation – in welchem Maße auch immer – in ihrem Nachschlagewerk zu dokumentieren, welche Strukturen der Informationsdarstellung erweisen sich dann als die optimalen, die den Zielvorstellungen bzw. Ansprüchen des Werks und seiner NutzerInnen gerecht werden?

Das Ziel des Beitrags kann und wird es nicht sein, erschöpfende Antworten auf diese Fragen zu geben, sondern vielmehr die Komplexität und Tragweite der angedeuteten Problemfelder zu diskutieren, und dies – wie oben erwähnt – im

Kontext standardsprachorientierter Lexikographie. Diesem Fokus gemäß wird zunächst in Kap. 2 das Verhältnis von Standardsprache und Variation aus theoretischer Perspektive beleuchtet, bevor in Kap. 3 konkrete Beispiele lexikographischer Variationsdokumentation in standardsprachlichen Wörterbüchern herangezogen werden. Auf Basis eines konkreten Variationsphänomens wird dann (in Kap. 4) die Vielfalt lexikalischer Variation in auch standardschriftsprachlichen Kontexten aufgedeckt und Möglichkeiten der lexikographischen Erfassung dieser Vielfalt diskutiert. Der Beitrag schließt mit einer Zusammenfassung und einem Ausblick, in dessen Rahmen ein Plädoyer für eine variationssensitive Lexikographie formuliert wird (Kap. 5).

## **2. Zum Verhältnis von Standardsprache und Sprachvariation – Theoretische Vorüberlegungen**

„German is probably the most varied language in Europe.“ (Barbour/Stevenson 1998: 2) – Die für das Deutsche nachweislich vorhandene Variationsvielfalt bedeutet für germanistische VariationslinguistInnen einen Glücksfall. Für LexikographInnen kann sie einen Problemfall darstellen, der sich – gerade im Hinblick auf standardsprachliche Lexikographie – durch jüngste Tendenzen *innerhalb* der Sprachrepertoires von Sprechenden – nicht nur des Deutschen, sondern in Europa generell – im 20. und 21. Jahrhundert verschärft. Wie Auer (2018) skizziert, beobachten wir in vielen Sprachgemeinschaften Europas einen Trend von ehemals diglossischen oder diaglossischen Dialekt/Standard-Repertoires hin zu monoglossischen. Das bedeutet, dass an die Stelle eines intraindividuellen Nebeneinanders von Standard und Dialekt im Sprachrepertoire von Sprechenden („innere Mehrsprachigkeit“) zunehmend lediglich „stilistische“ Variation innerhalb eines monoglossischen, in der Regel nur mehr ausschließlich standardsprach(nahen) Repertoires tritt (neben auch zunehmend „äußerer Mehrsprachigkeit“, wie sie etwa im Nebeneinander von Englisch und Deutsch vorliegt). Mit dieser Entwicklung hin zur intralingualen (standardsprachlichen/-nahen) Monoglossie gehen aber auch andererseits Prozesse der sozio-symbolischen Auf- bzw. Umbewertung von Varianten innerhalb der Standardsprache einher, die zunehmend als Marker regionaler, nationaler, sozialer (u. a.) Identität fungieren und damit Funktionen übernehmen, die ehemals auch bzw. vor allem nonstandardsprachlichen Varianten (insbesondere den Dialekten) zukamen.

Die sozio-symbolische Bedeutung sprachlicher Vielfalt *innerhalb* der deutschen Standardsprache wird gerade im Kontext „plurizentrischer“ Forschung deutlich. Unter den Termini *Plurizentrik/Plurizentrität* wird in der Forschungsliteratur die Tatsache gefasst, dass Deutsch in verschiedenen Ländern mit zumindest teilweise divergierenden standardsprachlichen Normen und Kodifizierungen vorzufinden ist (s. Clyne 1989; zur Plurizentrik-Diskussion im Hinblick auf den deutschen Sprachraum s. Schmidlin 2011, Lenz 2014, Herrgen 2015, Koppensteiner/Lenz 2017, Lanwermeier [et al.] 2019 und Koppensteiner/Lenz [eingereicht]). Nach Ammon/Bickel/Lenz (2016: 39) sind sogenannte „Zentren“ einer plurizentrischen Sprache „diejenigen Länder oder Regionen, die eigene standardsprachliche Besonderheiten herausgebildet haben. Von einem ‚Vollzentrum‘ spricht man dann, wenn die standardsprachlichen Besonderheiten in eigenen Nachschlagewerken, vor allem Wörterbüchern, festgehalten und autorisiert sind.“ Da diese Bedingung – zumindest im Hinblick auf lexikalische Normen – für Österreich (s. etwa ÖWB 2018 oder Ebner 2019), Deutschland (s. etwa Duden-Universalwörterbuch 2015) und die Schweiz (s. etwa Bickel/Landolt 2018) gegeben sind, werden diese drei Länder als „nationale Vollzentren der deutschen Sprache“ angesehen (Ammon/Bickel/Lenz 2016: 39). Inwieweit das Vorhandensein standardsprachlicher Normen mit der Existenz eines Kodexes gleichzusetzen ist, ist ebenso umstritten (s. Scharloth 2005) wie die Frage, ob das Deutsche als „plurinationale“ oder aber als „pluriareale“ Sprache zu klassifizieren ist (zur Diskussion s. auch Scheuringer 1996, Glauning 2013, Koppensteiner/Lenz 2017, Lenz/Koppensteiner [eingereicht]). Im Gegensatz zu den oben exemplarisch genannten (vor allem) landesspezifisch ausgerichteten Nachschlagewerken kommt im „Variantenwörterbuch“ (zur neuesten Auflage s. Ammon/Bickel/Lenz 2016) explizit – und insbesondere durch die Korpuszusammensetzung – ein pluriarealer Ansatz zum Einsatz, der der Tatsache Rechnung trägt, dass sich lexikalische Variation (auch *innerhalb* standardschriftsprachlicher Modelltexte (s. etwa Zeitungstexte) häufig nicht an Landesgrenzen orientiert, sondern erstens auch intranational lexikalische Variation zu beobachten ist und zweitens auch landesgrenzenüberschreitend lexikalische Gemeinsamkeiten aufzufinden sind, die sich auf Teilregionen mehrerer Länder verteilen.

In der Diskussion um das standardsprachliche Verhältnis von Deutschland, Österreich und der Schweiz ist häufig von einem „asymmetrischen“ Verhältnis der drei Zentren die Rede, das darin zum Ausdruck komme, dass das bundesdeutsche Deutsch in verschiedener Hinsicht die anderen beiden Zentren bzw. ihre standardsprachlichen Varietäten dominiere (s. etwa Schmidlin 2011, Wiesinger 2015). Die Wahrnehmung dieser einseitigen Beeinflussung kommt auch in Analysen von

Spracheinstellungen und Sprachperzeption zum Ausdruck (s. Koppensteiner/Lenz 2017, Koppensteiner/Lenz [eingereicht]), etwa wenn es um die Bewertung der „Standardsprachlichkeit“ oder „Korrektheit“ von nationalen Varianten geht (s. etwa Ammon 1995, Scharloth 2005, Schmidlin 2011).

Die vorangehenden Überlegungen verdeutlichen, dass vor Beantwortung der in Kap. 1 formulierten Fragen zum Verhältnis von Variation und (standardsprachorientierter) Lexikographie zunächst die vorgelagerte und mehr grundsätzliche Frage nach dem konkreten Standardsprachkonzept zu diskutieren ist, das einem lexikographischen Nachschlagewerk zugrunde gelegt wird. Welche Rolle nimmt im jeweiligen Standardsprachkonzept sprachliche Variation ein? (Siehe hierzu auch die Diskussion im IDS-Jahresband Eichinger/Kallmeyer 2005). Zugespitzt konkurrieren hier – in der Variationslinguistik vielleicht noch stärker als in der Lexikographie – zwei extreme Standpunkte, die einerseits Standardsprache als überregionales und variationsarmes oder sogar variationsfreies Konstrukt abheben von Standardsprachkonzepten andererseits, die neben fließenden Übergängen zum Nonstandard auch von areal-horizontaler Variation *innerhalb* des Standards ausgehen (zur Diskussion s. Deppermann/Kleiner/Knöbl 2013, Klein 2013, Kleiner 2014, Herrgen 2015, Lanwermyer [et al.] 2019).

### **3. Lexika und Variation – Analysen aus dem deutschsprachigen Raum**

#### **3.1 Allgemeine Beobachtungen und ihre Implikationen**

Lexikographische Nachschlagewerke, die nicht ausschließlich normativ-präskriptiv, sondern (auch oder verstärkt) empirisch-deskriptiv vorgehen, stehen grundsätzlich vor der Herausforderung, sprachliche Einheiten (i. d. R. „Wörter“ oder „Phrasen“) in ihrer linguistischen wie sozio-pragmatischen Komplexität und Vielfalt zu dokumentieren. Zu einer vollständigen linguistischen Beschreibung von lexikalischen Einheiten gehören Informationen zur Artikulation, Orthographie und Morphologie (flexionsmorphologische Aspekte sowie Wortbildungsmöglichkeiten), zur syntaktischen Einbettung in Satzkontexte sowie natürlich insbesondere zur semantischen Vielfalt, die optimalerweise mit Bezug auf den jeweiligen syntaktischen Kontext, in der eine bestimmte Lesart evoziert wird, angegeben wird. Die linguistischen Informationen müssen ergänzt werden durch Hinweise zu sozio-pragmatischen Steuerungsfaktoren. Diese Hinweise werden in der Regel nicht einem gesamten Lemma-Eintrag als Attribut zur Seite gestellt (wie dies im Ausnahmefall von *kriegen* in Kap. 3.2 der Fall ist), sondern sie beziehen sich häufiger nur auf

bestimmte Teilaspekte des Artikels und markieren damit nur bestimmte phonetische, morphologische, syntaktische oder semantische Ausprägungen der lexikalischen Einheit als von der unmarkierten Norm „abweichend“. Zu sozio-pragmatischen Steuerungsfaktoren, die in (exemplarisch ausgewählten) lexikographischen Nachschlagewerken der Gegenwart Verwendung finden, gehören etwa die folgenden Meta-Kommentierungen, die auf verschiedene Variationsdimensionen hinweisen (herangezogen wurden exemplarisch: Duden-Universalwörterbuch 2015, die Erstaufgabe des Variantenwörterbuchs (Ammon/Bickel/Ebner 2004), DWDS, Duden-Zweifelsfälle 2016, ÖWB 2018, Bickel 2018, Ebner 2019):

- **Areal-horizontale Variation:** Hinweise auf eine sprachgeographische Einschränkung einer Variante liefern Attribute wie „landschaftlich“ oder „regional“ bzw. spezifischer auf die konkrete räumliche bzw. nationale Distribution hinweisend Markierungen wie „schweizerisch“, „süddeutsch“ oder „österreichisch“. Diese Markierungen werden in der Regel nicht eingesetzt, um denotativ auf die Verbreitung einer standardsprachlichen Variante hinzuweisen, die lediglich regional spezifisch auftritt (so etwa im Variantenwörterbuch, s. Ammon/Bickel/Lenz 2016), sondern um auf eine Abweichung von der „eigentlichen“ (meist norddeutsch ausgerichteten) Standardnorm hinzuweisen, die sich – in nicht weiter konkretisiertem Ausmaß bzw. ohne Kontexteinbettung – als „weniger richtig“ oder aber zumindest als stärker markiert erweist.
- **Mediale Variation:** Auf verschiedene Grade konzeptioneller wie medialer Mündlichkeit bzw. Schriftlichkeit („Nähe-/Distanzsprache“) referieren Angaben wie „mündlich“ oder „Ausruf“. (Letztere verweist natürlich nicht nur auf die Verankerung der Variante in der Mündlichkeit, sondern enthält ebenso Hinweise auf systemlinguistische Aspekte wie etwa Suprasegmentalia, Wortartenzugehörigkeit u. a.)
- **Stilistische Variation:** Hinweise auf diese Variationsdimension liefern Meta-vokabeln wie „salopp“, „gehoben“, „dichterisch“ oder „derb“.
- **Soziolektale Variation:** Attribuierungen wie „Jugendsprache“/„jugendsprachlich“ oder „Gaunersprache“/„gaunersprachlich“ referieren auf gesellschaftliche Gruppen, die als prototypische SprecherInnen (im Kollektiv) der in Rede stehenden Varianten angesehen werden.

- **Funktiolektale Variation:** Dass Wörter v. a. in bestimmten Fachkontexten und in deren (meist auch normierter und kodifizierter) Terminologie auftreten, deuten Zusätze wie „fachsprachlich“ bzw. konkreter „Mathematik“, „Sprachwissenschaft“ oder „Ökonomie“ an.
- **Pragmatische Variation:** Auf pragmatische Aspekte verweisen Attribute wie „spöttisch“, „abwertend“, „vertraulich“, „ironisch“, „vulgär“ oder „emotional“. Sie informieren über die (potentielle) Adäquatheit einer Variante (in bestimmten sozio-pragmatischen Kontexten), die soziale Beziehung, die zwischen den KommunikationsteilnehmerInnen durch die Verwendung der Variante besteht bzw. verändert werden kann, die Wirkung, die ihre Verwendung auf RezipientInnenseite (womöglich) auslöst, sowie auf die psychosoziale Verfasstheit der Sprechenden bzw. die individuelle Relation/Haltung zum perspektivierten Gegenstand (s. „ironisch“).
- **Frequentielle bzw. diachrone Variation:** Hinweise zur Häufigkeit bzw. dem Alter einer Variante verbergen sich hinter Attributen wie „selten“ oder „veraltet“, die auf Gebrauchsfrequenzen bzw. den Aktualitätsgrad und damit auch die Wahrscheinlichkeit des Auftretens innerhalb des Gegenwartssprachlichen verweisen. Mitunter, aber nicht zwingend, kann aus diesen Frequenzen auf das „Alter“ einer sprachlichen Variante geschlossen werden (und vice versa), insofern als „seltene“ Varianten häufig verschwindende sprachliche Entitäten darstellen bzw. „veraltete“ Varianten auch nur mehr in geringen Gebrauchsfrequenzen auftreten.

### Und was ist mit „umgangssprachlicher“ Variation?

Während die zuvor exemplarisch herangezogenen Markierungen lexikographische Zusatzinformationen liefern, die meist recht eindeutig einer der genannten sozio-pragmatischen Variationsdimensionen zugeordnet werden können, handelt es sich bei der womöglich am häufigsten in gegenwartssprachlichen Wörterbüchern zum Deutschen eingesetzten Meta-Vokabel „umgangssprachlich“ um eine Angabe, deren semantisch-funktionale Entschlüsselung sich nicht selten als schwierig erweist. Die Vieldeutigkeit, die einerseits die hohen Gebrauchsfrequenzen des Attributs, aber eben auch seine Problematik (nicht nur) innerhalb lexikographischer Kontexte ausmacht, wird etwa in folgendem Definitionsversuch des Duden-Universalwörterbuch (2015: 18f.) deutlich:

Die mit der Abkürzung „ugs“. markierte »Umgangssprache« ist der Sprachstil, wie man ihn im alltäglichen „Umgang“ zwischen den Menschen – vor allem – hört bzw. dort liest, wo individuelle Abweichungen von der Norm der Hochsprache üblich sind, etwa in persönlichen Briefen und mittlerweile auch in Chatroom- und E-Mail-Texten. [...] Die rasante Entwicklung auf dem Gebiet der Massenmedien während der letzten zwanzig Jahre hat unsere Gesellschaft auf den Weg zu einer Multimediagesellschaft gebracht, die prägend in den täglichen Umgang der Menschen miteinander eingreift und damit auch auf die »Umgangssprache« ausstrahlt. [...] So gesehen, beginnt die Markierung „ugs.“ zunehmend die negative Aura zu verlieren, die ihr traditionell in Wörterbüchern anhaftet. Sie beschreibt eher einen Umgang mit der Sprache, der zwar von der Sprachnorm abweicht, jedoch eine hohe Frequenz wie auch eine breite Akzeptanz aufweist.

Unabhängig von der begrifflichen Fassung und Operationalisierbarkeit der Kategorie „umgangssprachlich“ bezieht sie sich – wie auch die anderen metasprachlichen Angaben – in der Regel nicht auf einen Lexikoneintrag als Ganzes, sondern lediglich auf bestimmte Ausschnitte des Eintrags und damit auf bestimmte Aussprachevarianten, ausgewählte morphologische Paradigmenformen, bestimmte syntaktische Kontexte, in denen eine Lexikoneinheit eingebettet ist, oder auf bestimmte Bedeutungsvarianten. In der Regel treten also linguistische und sozio-pragmatische Steuerungsfaktoren einer sprachlichen Einheit in multidimensionaler Verflechtung auf.

### 3.2 Lexikographische Fallbeispiele in der Tiefenanalyse

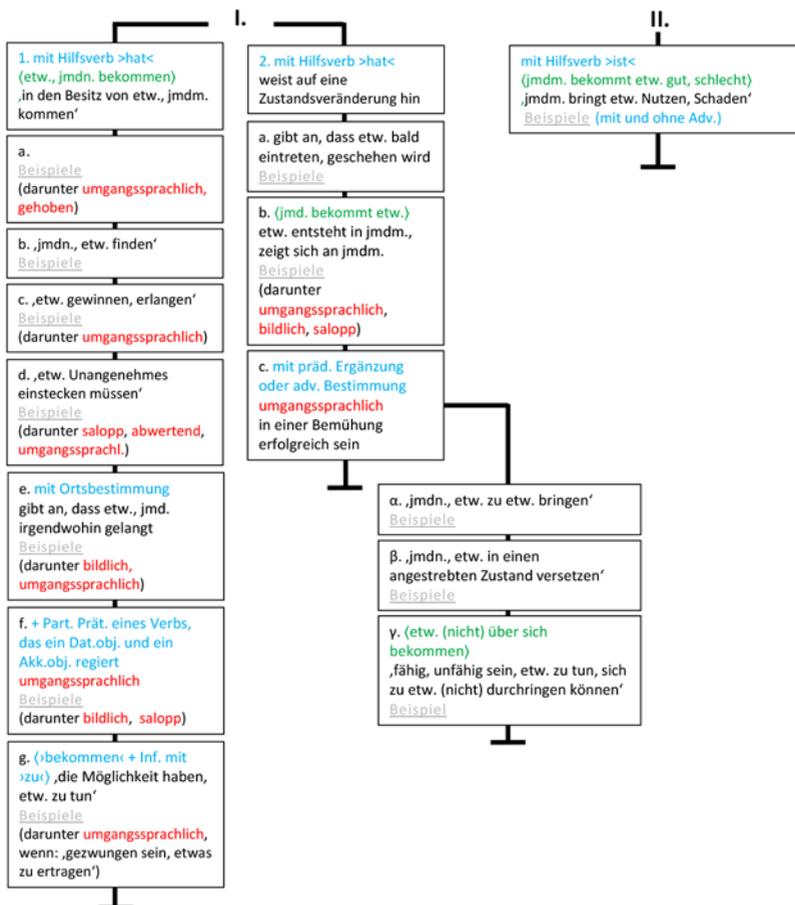
Wie geht ein und dasselbe Nachschlagewerk konkret mit den zuvor skizzierten linguistischen wie soziolinguistischen Variationsdimensionen und mit diesen in ihren Verflechtungen um? Um diese Frage zu beantworten, werden im Folgenden die Lexikonartikel zweier Quasisynonyme aus dem DWDS (<https://www.dwds.de>) herangezogen, nämlich die (insgesamt drei) Artikel der Verben *kriegen* (aufgeteilt in *kriegen*<sup>1</sup> und *kriegen*<sup>2</sup>) und *bekommen*, deren Auswahl durch mehrere Besonderheiten motiviert ist (s. ausführlicher Lenz 2013: 5–45): Es handelt sich um hochfrequente Verben der deutschen Sprache (s. auch die Worthäufigkeitsangaben im DWDS), die sich durch eine hohe semantische wie syntaktische Komplexität ausweisen. Die Gebrauchspräferenzen bzw. der standardsprachliche Status ihrer diversen Bedeutungs- bzw. Konstruktionsvarianten gehen mit komplexen

linguistischen wie soziopragmatischen Steuerungsfaktoren einher. Ein kontrastiver Blick auf beide Verben wird durch ihr besonderes „Konkurrenzverhältnis“ motiviert, das auch mit affektiv-evaluativen Ladungen versehen ist (s. hierzu auch die gesammelten Ausschnitte aus Internetforen in Lenz 2013: 403–417). Die im Folgenden auf Basis der DWDS-Artikel herausgearbeiteten Generalisierungen stimmen in hohem Maße mit anderen (standardsprachorientierten) Nachschlagewerken zum Gegenwartsdeutschen (z. B. Duden-Universalwörterbuch 2015) überein. Der Fokus der folgenden Diskussion wird auf den syntakto-semantischen Angaben sowie den metasprachlichen Markierungen der Lexikoneinträge liegen.

**DWDS-Artikel *bekommen*:** <https://www.dwds.de/wb/bekommen>  
[letzter Zugriff: 10.10.2019]

Die Produktivität des Verbs *bekommen* im Rahmen von Wortbildungsprozessen wird bereits durch die Anzahl an Präfigierungen, die das DWDS für „*bekommen*‘ als Letztglied“ aufführt, deutlich, nämlich insgesamt 32 (z. B. *abbekommen*, *durchbekommen*, *mitbekommen*). Auch das DWDS-Feld „Bedeutungen“, das aus dem WDG (1967, Band 1) übernommen wurde, weist allein durch die Anzahl von (auf unterster Strukturebene) 14 differenzierten Bedeutungsvarianten auf die semantische Komplexität des Verbs hin (s. Abb. 1). Gegliedert sind diese Bedeutungsvarianten auf einer ersten Ordnungsebene hinsichtlich ihres Perfektauxiliars: Bis auf eine Ausnahme sind alle Varianten der Rubrik „mit Hilfsverb ‚hat‘“ zugeordnet (s. „I.“ in Abb. 1), lediglich die Bedeutung „jmdm. bringt etw. Nutzen, Schaden“ repräsentiert die Rubrik „mit Hilfsverb ‚ist‘“ (s. „II.“ in Abb. 1). Unter den Varianten mit Perfektauxiliar *haben* heben sich Subvarianten mit der Bedeutung „in den Besitz von etw., jmdm. kommen“ ab von Subvarianten, die „auf eine Zustandsveränderung hin[weisen]“. Wie die Blaufärbung in Abb. 1 verdeutlichen will, sind nur einige der 14 Subvarianten auch mit syntakto-semantischen Informationen versehen, die Hinweise zum konkreten Konstruktionskontext enthalten, mit dem eine bestimmte Bedeutung verknüpft ist. Die Art und Weise, wie diese Informationen vermittelt werden, variiert im Artikel: Mitunter wird die Konstruktion unter Rückgriff auf grammatische Terminologie syntaktisch definiert, wie es im Falle der Variante 2f der Fall ist („+ Part. Prät. eines Verbs, das ein Dat.obj. und ein Akk.obj. regiert“), ohne dass weitere Hinweise auf die Semantik dieser Variante/Konstruktion gegeben werden. In anderen Fällen wird die Semantik einer Variante paraphrasiert, ohne dass grammatische (oder sonstige) Zusatzinformationen explizit geliefert werden. Dies ist etwa der Fall bei den Bedeutungsvarianten 1b bis 1d. Indirekte Informationen zum

Valenzrahmen des Verbs verstecken sich jedoch in den Paraphrasierungen dieser Bedeutungsvarianten, indem Angaben wie „etw.“, „jmdn.“ oder „jmdm.“ auf die syntaktischen Anschlussfähigkeiten des Verbs hindeuten, das mit direktem bzw. indirektem Objekt verknüpft werden kann, welches auf animate („jmdn.“ bzw. „jmdm.“) oder inanimate („etw.“) Entitäten referiert. In wiederum anderen Fällen wird eine Kombination aus semantischer Paraphrasierung und grammatischer Konstruktionsbeschreibung gewählt, wie es etwa bei der Variante 1g der Fall ist: „bekommen‘ + Inf. mit ‚zu‘ ‚die Möglichkeit haben, etw. zu tun‘“.



**Abb. 1:** Struktur (schematisiert) des DWDS-Artikels *bekommen* (<https://www.dwds.de/wb/bekommen> [letzter Zugriff: 28.09.2019])

Ähnlich komplex wie die linguistische Vielfalt des Verbs *bekommen* gestaltet sich seine sozio-pragmatische Variation, die sich in Markierungen widerspiegelt, die in Abb. 1 in rot wiedergegeben werden (grün im DWDS-Original). Mitunter ist es nicht eine Konstruktion an sich, sondern es sind lediglich bestimmte lexikalische Füllungen derselben, die als „normabweichend“ markiert werden. Als Beispiel dient die Variante 1e „mit Ortsbestimmung“: Während *er hat dieses Buch versehentlich in die Hände bekommen* ein unmarkiertes Beispiel darstellt, wird *er hat dieses Buch versehentlich in die Finger bekommen* als „umgangssprachlich“ eingestuft. Anders gelagert ist hingegen der Fall des Dativpassivs mit dem Auxiliär *bekommen* (s. Variante 1f), das per se als „umgangssprachlich“ eingestuft wird (s. aber Kap. 4). Verkomplizierend kommt hinzu, dass die lexikalisierte Phrase *etw. aufs Brot geschmiert bekommen* zusätzlich als „bildlich“ und „salopp“ ausgewiesen wird. Besonders komplex gestalten sich die metasprachlichen Angaben im Hinblick auf Variante 1d, bei der fast sämtliche Belegbeispiele mit einer Kombination von bis zu drei semantisch nah beieinanderliegenden sozio-pragmatischen Labels versehen werden (s. Abb. 2).

d) *etw. Unangenehmes einstecken müssen*

BEISPIELE:

eine Rüge, Lehre **bekommen**

Prügel, eine Ohrfeige **bekommen**

salopp, abwertend einen Anschauzer **bekommen**

umgangssprachlich, salopp, abwertend etw. hinter die Ohren **bekommen**

salopp, abwertend er hat sieben Monate (Gefängnis) **bekommen**

umgangssprachlich einen Dämpfer **bekommen**

salopp, umgangssprachlich Dresche, sein Fett **bekommen**

umgangssprachlich er hat eins aufs Dach **bekommen**

Abb. 2: Variante 1d aus dem DWDS-Artikel *bekommen* (<https://www.dwds.de/wb/bekommen> [letzter Zugriff: 10.10.2019])

DWDS-Artikel *kriegen*<sup>1</sup> und *kriegen*<sup>2</sup>: <https://www.dwds.de/wb/kriegen#1> [letzter Zugriff: 10.10.2019] und <https://www.dwds.de/wb/kriegen#2> [letzter Zugriff: 10.10.2019]

Ein anderes Bild ergibt sich hingegen für das (transitive) Verb „*kriegen*<sup>1</sup>“, das unabhängig von seinem lexikalischen oder syntaktischen Kontext – ebenso wie seine 41 Wortbildungen „mit ‚*kriegen*‘ als Letztglied“ – per se als „umgangssprachlich“

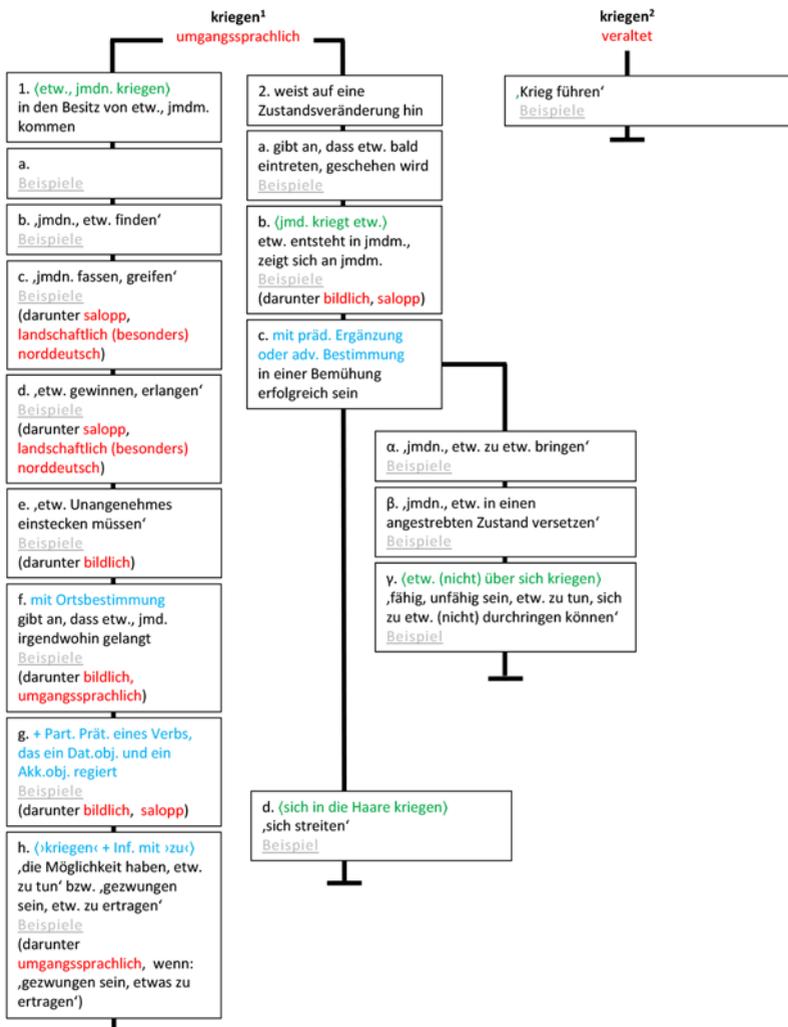


Abb. 3: Struktur (schematisiert) der DWDS-Artikel *kriegen*<sup>1</sup> (<https://www.dwds.de/wb/kriegen#1>) und *kriegen*<sup>2</sup> (<https://www.dwds.de/wb/kriegen#2> [letzter Zugriff: 28.09.2019])

markiert wird (s. auch Abb. 3). Trotz dieser generalisierenden Attribuierung des Verbs an sich und damit all seiner Varianten finden sich mitunter weitere sozio-pragmatische Markierungen (s. Abb. 3), wie sie etwa ausgewählten Phrasen der Variante 1c zur Seite gestellt werden, die als „salopp; landschaftlich, (besonders) norddeutsch“ ausgewiesen werden. Besonders irritierend erscheint die Stilmarkierung zu „ein paar hinter die Ohren kriegen“ (s. 1e), weil hier das Attribut „umgangssprachlich“ noch einmal wiederholt auftritt.

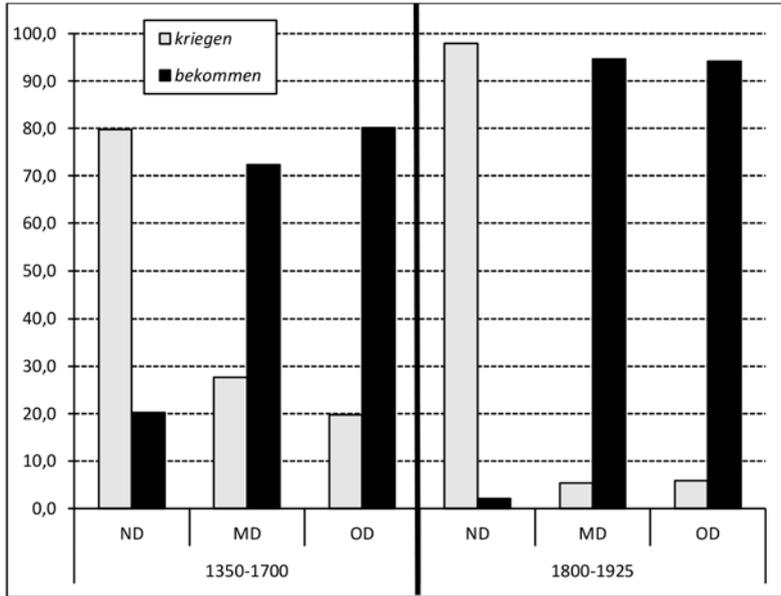
Im Gegensatz zu den transitiven *kriegen*-Varianten, die alle unter *kriegen*<sup>1</sup> zusammengefasst werden, wird intransitives *kriegen* in der Semantik von „Krieg führen“ als eigener Lemmaeintrag unter *kriegen*<sup>2</sup> gelistet, der nicht als „umgangssprachlich“, sondern als „veraltet“ klassifiziert wird. Wie korpuslinguistische Analysen in Kap. 4 zeigen werden, ist diese *kriegen*-Variante allerdings lediglich in Deutschland und Österreich eine archaische Variante, die in standardschriftsprachlichen Kontexten der Deutschschweiz aber nach wie vor auftritt.

#### 4. Korpusanalytische Befunde (nach Lenz 2013)

Den lexikographischen Informationen, die zuvor aus den DWDS-Einträgen der Verben *kriegen* und *bekommen* herausgearbeitet wurden, werden in einem nächsten Schritt korpuslinguistische Befunde gegenübergestellt, die ausführlicher bereits in Lenz 2013 präsentiert sind. Im Folgenden werden lediglich die für die lexikographische Diskussion interessanten Ergebnisausschnitte skizziert. Die herangezogenen Befunde basieren auf umfangreichen diachronen wie gegenwartsbezogenen Korpusanalysen, die nicht nur den standardsprachlichen Bereich der „vertikalen“ Variationsdimension betreffen, sondern das Gesamt arealsprachlicher Varietäten der Dialekt-Standard-Achse. Diese umfassende Perspektivierung trägt der Überzeugung Rechnung, dass nur die varietäre Zusammenschau der gesamten vertikalen Achse von Dialekt bis Standard die notwendigen Informationen liefert, um das komplexe Zusammenspiel der konkurrierenden Varianten von *kriegen* und *bekommen* im deutschsprachigen Raum fundiert greifen zu können. Neben der areal-horizontalen und vertikal-sozialen Variationsdimension bedarf es aber zusätzlich – und auch darauf deuten zahlreiche sozio-pragmatische Attribuierungen in den Lexikoneinträgen hin (s. Kap. 3) – der Berücksichtigung einer zeitlichen und medial-konzeptionellen Perspektivierung der Verben.

Die historischen Korpora, die in Lenz (2013: Tab. 25) zur Analyse der syntaktosemantischen Variation von *kriegen* und *bekommen* herangezogen wurden, decken den hochdeutschen wie niederdeutschen Raum in der Zeit von 1350 bis 1925 ab. Ihnen zur Seite gestellt sind (mehr) gegenwartsbezogene Korpora bestehend aus mündlichen und geschriebenen Daten, die ebenfalls regional gestreut sind, indem sie die sechs sprachlichen Großregionen des deutschsprachigen Raums abdecken (Nord-, Mittel- und Oberdeutsch in ihrer jeweiligen Ost- und Westausprägung). Die dialektalen Analysen basieren zum einen auf den großlandschaftlichen und anderen Dialektwörterbüchern des gesamten deutschen Sprachraums (s. Übersicht in Niebaum/Macha2 2006: 37) und zum anderen auf den Tonaufnahmen aus dem „Zwirner-Korpus“. (Beim „Zwirner-Korpus“ handelt es sich um ein Korpus bestehend aus mehr als 5.800 Dialektaufnahmen, die zwischen 1950 und 1960 unter der Leitung von Eberhard Zwirner an ca. 1.000 ländlichen Ortspunkten in den alten Bundesländern, in Vorarlberg und Liechtenstein, im Elsass und in den Niederlanden gesammelt wurden. Die Aufnahmen sowie die Transkriptionen von ca. 3.000 Aufnahmen sind über die „Datenbank Gesprochenes Deutsch (DGD)“ des Instituts für deutsche Sprache (IDS) verfügbar (<https://dgd.ids-mannheim.de/>). Die hier analysierten Daten basieren auf den Zwirner-Aufnahmen folgender „Mundartgruppen“: Holsteinisch, Ostpommersch, Mittelfränkisch, Schlesisch, Obersächsisch, Nieder- und Hochalemannisch sowie Mittelbairisch.) Den Zwirner-Daten sind die „umgangssprachlichen“ Aufnahmen des „Pfeffer-Korpus“ sowie Karten aus Regiolekt-Atlanten (z. B. AdA) gegenübergestellt. (Das „Pfeffer-Korpus“ umfasst Aufnahmen von insgesamt 403 städtischen „UmgangssprachesprecherInnen“ aus insgesamt 57 Städten der alten Bundesrepublik, der ehemaligen DDR, Österreichs und der Schweiz.) Einblicke in die schriftsprachlichen Verhältnisse liefern einerseits Ausschnitte aus IRC-Städtechats (IRC = „Internet Relay Chat“; bei den ausgewerteten Daten handelt es sich um Ausschnitte aus IRC-Gesprächskanälen nord-, mittel- und süddeutscher Städte, die vor allem zwischen Januar und September 2006 gesammelt wurden) sowie andererseits Zeitungstexte aus dem „Deutschen Referenzkorpus (DeReKo)“ des Instituts für deutsche Sprache (IDS) (<http://www1.ids-mannheim.de/kl/projekte/korpora.html>), die den divergierenden Graden konzeptioneller Mündlichkeit in schriftlichen Daten Rechnung tragen.

Abb. 4 kontrastiert die relativen Häufigkeiten, in denen die beiden Verben in den Zeiträumen 1350 bis 1700 bzw. 1800 bis 1925 jeweils im Nieder-, Mittel- und Oberdeutschen auftreten, wobei den beiden Verben absolute Häufigkeiten von 657 (*kriegen*) bzw. 577 (*bekommen*) Belegen in den historischen Korpora



**Abb. 4:** Relative Häufigkeiten (in %) von *kriegen* (n = 657) und *bekommen* (n = 577) im Nieder-, Mittel- und Oberdeutschen im Vergleich der beiden Zeitschnitte (Lenz 2013: Abb. 65)

zugrunde liegen. Das Diagramm verdeutlicht zwei entgegengesetzte Prozesse, die den niederdeutschen Sprachraum auf der einen Seite und den hochdeutschen Sprachraum auf der anderen Seite voneinander abheben. Während *bekommen* im Mittelniederdeutschen (s. erster Zeitschnitt) als Minderheitenvariante etwa noch ein Fünftel der Belege ausmacht, ist es zum Neuniederdeutschen hin zugunsten von eindeutig dominierendem *kriegen* fast vollständig abgebaut. Im Frühneuhochdeutschen hingegen ist es *kriegen*, das noch etwa 20 bis 30 % der gemeinsamen Belege besetzt, bevor es im zweiten Zeitraum zugunsten von *bekommen* zum seltenen Ausnahmefall wird (zumindest in den hier analysierten schriftsprachlichen Daten). Wie der Vergleich zeigt, ist somit seit Jahrhunderten von einer Koexistenz der beiden Verben *bekommen* und *kriegen* im deutschen Sprachraum auszugehen, mit arealen Schwerpunkten von *kriegen* im Norden und von *bekommen* im Süden. Im hochdeutschen Sprachraum hat die „Vertikalisierung“ (Reichmann [u. a.] 1988) der ehemals primär areal-horizontal distribuierten Varianten eine mittel- und oberdeutsch quantitativ dominierende Variante zu einer auch sozial höherstehenden Prestigevariante werden lassen (s. ausführlicher Lenz 2013: 417–427).

Lesart (Typ)	Syntaktisches Muster	kriegen	bekommen	Paraphrasierungen
COME (emerge')	intrans. V		+	'heraustreten', 'wachsen', 'gedehen', 'befruchtet sein'
COME (agree')	zich V (teziprok)		+	'miteinander übereinkommen', 'sich einig werden'
COME (get by')	intrans. V + NP, AdvP		+	'mit etwas hin- / auskommen'
COME (meet')	intrans. V + NP, AdvP		+	'jemandem entgegenkommen / begegnen'
COME	V + NP, AdvP (+ PP, AdvP) PartikelP		+	'einen Ort erreichen, an einen Ort kommen', 'jemandem antreffen / einholen'
COME	intrans. V + (AdvP) PartikelP <sub>dir</sub>		+	'sich rumbewegen (wegbewegen)'
BENEFICIAL	intrans. V (+ NP, AdvP)		+	
FIGHT	intrans. V (+ PP)		+	'Krieg führen', 'streiten', 'schimpfen', 'sich abmühen'
FIGHT	V + NP, AdvP (+ PP, AdvP) PartikelP		+	'bekämpfen', 'auschüpfen', 'zerstören'
FIGHT	zich V (Partikel) V (+ PP)		+	'sich streiten', 'sich herumschlagen', 'sich abfinden'
'reproach'	V + NP, AdvP + NP, AdvP		+	'jdm. NP, AdvP vorhalten'
GET (= EXERT) RECEIVE	V + NP, GEN		+	'NP (mit ohne eigenes Zutun) in seine Kontrolldomäne bekommen'
GET (= EXERT)	V + NP, AdvP (+ PP, AdvP) PartikelP		+	'schaffen, NP in seine Kontrolldomäne zu bringen'
find'	zich V (teziprok)		(+)	'einander (als Partner) gewinnen'
RECEIVE	V + NP, AdvP (+ PP, AdvP) PartikelP		+	'NP ohne eigenes Zutun in seine Kontrolldomäne bekommen'
GET/TAKE	V + NP, AdvP + NP, AdvP		+	'jdm. NP, AdvP geben / wegnehmen'
TAKE GRASP	V + NP, AdvP (+ PP, AdvP) PartikelP		+	'NP nehmen (anfassen / packen)'
TAKE GRASP	zich V + NP, AdvP (+ PP, AdvP) PartikelP		+	'(sich) NP nehmen / fassen / packen'
GRASP	zich V + PP, AdvP (teziprok)		+	'einander (an einem Körperteil) packen / fassen'
'set better'	zich V (Partikel) V (reflexiv)		+	'sich erholen', 'sich wieder fangen'
inchoativ	V + NP, AdvP (+ PP, AdvP) PartikelP		+	'NP (als Zustands- / Umstandsveränderung) erfahren oder erwarten'
inchoativ	V + (PP, AdvP) PartikelP <sub>dir</sub> + NP, AdvP		+	'erfahren, dass NP an einen Ort bzw. in einen Zustand gerät'
inchoativ	V + AdvP + NP, AdvP		+	'erfahren, dass NP eine bestimmte Eigenschaft annimmt'
inchoativ	V + <i>das</i> / <i>es</i> + <i>mir</i> , PP		+	'NP (als Zustands- / Umstandsveränderung) erfahren oder erwarten'
inchoativ	inchoativ V + AdvP		+	'eine bestimmte Eigenschaft annehmen'
ingressiv	V + <i>das</i> + substantiv. Inf.		+	'anfängen etw. zu tun'
ingressiv	V + <i>das</i> / <i>es</i> + <i>mit</i> / <i>PP</i> , subst. Inf.		+	'anfängen etw. zu tun'
modal	zich V + <i>du</i> (m) Inf. + (NP, AdvP)		+	'anfängen etw. zu tun'
modal	V + Part. II		+	'etw. tun können'
PASSIV	V + Part. II (+ NP, AdvP)		+	'erfahren, dass mit NP bzw. mit einem selbst etwas getan wird'
resultativ + EXERT	V + (PP, AdvP) PartikelP <sub>dir</sub> + NP, AdvP		+	'machen / schaffen, dass NP einen Ort bzw. Zustand erreicht'
resultativ + EXERT	V + AdvP + NP, AdvP		+	'schaffen, dass NP eine bestimmte Eigenschaft annimmt'
resultativ + EXERT	V + Part. II + NP, AdvP		+	'machen / schaffen, dass mit NP etwas getan wird'
kausativ + EXERT	V + PP, subst. Inf. / (PP) / <i>zu</i> - Inf. / <i>das</i> / <i>zu</i> + NP, AdvP		+	'schaffen, NP dazu zu bringen, etwas zu tun'
MASTER	V + NP, AdvP (+ PP, AdvP) PartikelP		+	'NP schaffen, meist'
PUR/FETCH	V + (PP, AdvP) PartikelP <sub>dir</sub> + NP, AdvP		+	'NP an einen Ort bewegen bzw. von einem Ort wegbewegen'

Abb. 5: Kriegen- und bekommen-Typen in den Varietäten des Gegenwartsdeutschen, sortiert nach syntaktischen Mustern (s. Lenz 2013: Tab. 30)

Den historischen Befunden stellt Abb.5 die Ergebnisse der gegenwartsbezogenen Korpusanalysen gegenüber, basierend auf den oben zu Beginn des Kapitels aufgeführten Korpusquellen, die die vertikale Dialekt-Standard-Achse in ihrer Gesamterstreckung abbilden. Der Übersicht ist zu entnehmen, in welchen syntakto-semantischen Kontexten *kriegen* und *bekommen* prinzipiell gegenwartssprachlich miteinander konkurrieren, unabhängig von den konkreten Varietäten oder Räumen. Auch wenn hier nicht auf die Detailergebnisse eingegangen werden kann, macht bereits ein erster Blick auf die Verteilung der grau gefärbten Zellen deutlich, dass *kriegen* und *bekommen* nur in einer Teilmenge der aufgeführten syntakto-semantischen Kontexte in den Varietäten des Gegenwartsdeutschen koexistieren, genauer: Nur in 16 der insgesamt 37 eruierten syntakto-semantischen Kontexten ist eine Kovariation belegt, während in den übrigen Kontexten nur eines der beiden Verben auftritt. Bereits diese erste Übersicht lässt die oben aus standardsprachlicher Perspektive formulierten metalinguistischen Markierungen in einem kritischen Licht erscheinen: *kriegen* ist einerseits mehr, andererseits weniger als nur das „umgangssprachliche“ Pendant zu *bekommen*, und *bekommen* ist einerseits weniger, andererseits mehr als nur das „hochsprachliche“ Pendant von *kriegen*.

Abb. 6 und Abb. 7 schränken die Beobachtungsperspektive ein, indem der Fokus nun auf Befunde gelenkt wird, die sich nicht auf alle in Lenz 2013 ausgewerteten gegenwartssprachlichen Korpora stützen, sondern lediglich auf Korpusanalysen zur Standardschriftsprache basierend auf ausgewählten Zeitungen im deutschsprachigen Raum. Neben drei Zeitungen aus dem Norden, der Mitte und dem Süden Deutschlands wurden das St.Galler Tagblatt (Schweiz) und die Niederösterreichischen Nachrichten aus dem Jahr 2009 analysiert. Auch wenn *bekommen* in den Zeitungskorpora etwas mehr als 90 % aller *kriegen/bekommen*-Belege ausmacht und damit – wie zu erwarten – die quantitativ dominierende Variante darstellt, nimmt *kriegen* immerhin knapp 10 % der Belege ein. *Bekommen* dominiert also die standardschriftsprachlichen Korpora bei weitem, aber auch *kriegen* kommt in den Zeitungstexten sehr wohl vor, und zwar in allen fünf differenzierten Teilkorpora (dabei im interregionalen Vergleich im St. Galler Tagblatt und in den Niederösterreichischen Nachrichten insgesamt seltener als in den drei Zeitungen aus Deutschland).

Qualitative Tiefenbohrungen zeigen, dass sich die beiden Verben zeitungsspezifisch und damit auch regionenübergreifend insgesamt auf recht ähnliche syntakto-semantische Kontexte verteilen (s. Lenz 2013: 459–471), was in den Abb. 6 und 7 im

Typ	Syntaktisches Muster	Lesart (Typ)	D			CH	A
			Hann Allg	RhZ	NüN	StGal	NieÖ
I.	intrans. <i>kriegen</i> (+ PP)	FIGHT				(x)	
II.	intrans. <i>kriegen</i> + AdjP	inchoativ					
III.	intrans. <i>kriegen</i> + (AdvP/PartikelP) <sub>dir</sub>	COME					
IV.	<i>kriegen</i> + NP <sub>GEN</sub>	GET/RECEIVE					
V.		FIGHT					
VI.		GET + EXERT	x	x	x	x	x
VII.	<i>kriegen</i> + NP <sub>AKK</sub>	TAKE/GRASP					
VIII.	(+ PP/AdvP/PartikelP)	MASTER	(x)	(x)	(x)	(x)	(x)
IX.		COME					
X.		RECEIVE	x	x	x	x	x
XI.		inchoativ	x	x	x	x	x
XII.	<i>kriegen</i> + NP <sub>AKK</sub> + NP <sub>DAT</sub>	FIGHT (reproach')					
XIII.		GIVE/TAKE (AWAY)					
XIV.	<i>kriegen</i> + <i>das</i> + substantiv. Inf.	ingressiv					
XV.	<i>kriegen</i> + <i>das/es</i> + mit-PP <sub>(subst. Inf.)</sub>	ingressiv					
XVI.	<i>kriegen</i> + <i>das/es</i> + mit-PP	inchoativ					
XVII.	<i>sich bekriegen</i>	FIGHT	x	x	x	x	x
XVIII.	<i>sich kriegen</i> (reziprok)	GET + EXERT (find')	x	x			
XIX.	<i>sich kriegen</i> + PP/AdvP (reziprok)	GRASP	(x)			(x)	(x)
XX.	<i>sich einkriegen</i>	FIGHT (.get better')	(x)				
XXI.	<i>sich kriegen</i> + NP <sub>AKK</sub> (+ PP/AdvP/PartikelP)	TAKE/GRASP					
XXII.	<i>sich kr. + das</i> + subst. Inf.	ingressiv					
XXIII.		resultativ + EXERT	x	x	x	x	x
XXIV.	<i>kr. + (PP/AdvP/PartikelP)<sub>dir</sub> + NP<sub>AKK</sub></i>	PUT/FETCH					
XXV.		inchoativ	x	x		x	x
XXVI.	<i>kriegen</i> + PP <sub>(subst. Inf./für) zu-Inf./dazu</sub> + NP <sub>AKK</sub>	kausativ + EXERT	x			(x)	
XXVII.		resultativ + EXERT	x	x	x	x	x
XXVIII.	<i>kriegen</i> + AdjP + NP <sub>AKK</sub>	inchoativ					
XIX.	<i>kriegen</i> + <i>zu(m)/für</i> zu-Inf. (+NP <sub>AKK</sub> )	modal	x	x	x	x	x
XXX.		resultativ + EXERT	x	x	(x)	(x)	
XXXI.	<i>kriegen</i> + Part. II (+ NP <sub>AKK</sub> )	modal					
XXXII.		passiv	x	x	x	x	x

Abb. 6: Areale Verteilung von *kriegen*-Belegen in den Zeitungskorpora, sortiert nach syntaktischen Mustern (nach Lenz 2013: Tab. 12)

Vergleich der grau hervorgehobenen Typen deutlich wird. Dennoch lassen sich einige interessante Detailbefunde beobachten. Diese betreffen etwa die bereits oben angesprochene FIGHT-Semantik von *kriegen*, die im DWDS in einen eigenen Lemma-Eintrag (*kriegen*<sup>2</sup>) gemündet ist. Es handelt sich um eine Lesart, die lediglich von *kriegen*, nicht aber von *bekommen* evoziert wird, die aber – anders als im DWDS-Eintrag aufgeführt – zumindest in der Schweiz keine veraltete Variante darstellt, wie der folgende Beleg aus dem St. Galler Tagblatt (19.10.2009: 29) zeigt: „Können wir die Verantwortung für so viele Tote weiterhin tragen? Nein, denn wenn die Schweiz nicht kriegen soll, dann soll sie auch nicht anderweitig am Krieg beteiligt sein.“ Ebenso wie eine „kriegerische“ Lesart lediglich von *kriegen* evoziert wird, stellt die Variante mit Dativargument in der Semantik „jemandem (gut, schlecht, irgendwie) zuträglich sein“ eine Besonderheit des Verbs *bekommen* dar, die nicht mit dem Verb *kriegen* assoziiert ist.

Eine weitere auffällige Differenz zwischen den syntakto-semantischen Kontexten betrifft die sogenannten „Resultativkonstruktionen“ mit exert-Semantik (s. Lenz

Typ	Syntaktisches Muster	Lesart	D			CH	Ö
			Hann Allg	RhZ	Nün	StGal	NieÖ
I.	intrans. <i>bekommen</i>	COME (,emerge')					
II.	intrans. <i>bekommen</i> + NP <sub>DAT</sub>	COME (,meet')					
III.	intrans. <i>bekommen</i> + mit-PP	COME (,get by')					
IV.	intrans. <i>bekommen</i> (+ NP <sub>DAT</sub> ) (+ AdvP)	BENEFICIAL	+	+		+	+
V.	<i>bekommen</i> + NP <sub>GEN</sub>	GET/RECEIVE					
VI.		COME (,meet')					
VII.		GET + EXERT	+	+	+	+	+
VIII.	<i>bekommen</i> + NP <sub>AKK</sub> (+ PP/AdvP/PartikelP)	MASTER					
IX.		RECEIVE	+	+	+	+	+
X.		inchoativ	+	+	+	+	+
XI.	<i>sich bekommen</i> (reziprok)	COME (,agree')					
XII.	<i>sich bekommen</i> (reziprok)	GET + EXERT (,find')					
XIII.	<i>sich bekommen</i> (reflexiv)	,get better'					
XIV.		resultativ + EXERT	+	+	+	+	+
XV.	<i>bekommen</i> + (PP/AdvP/PartikelP) <sub>adv</sub> + NP <sub>AKK</sub>	inchoativ	+	+	+	+	
XVI.	<i>bekommen</i> + PP <sub>subst. Inf</sub> / <i>dazu</i> + NP <sub>AKK</sub>	kausativ + EXERT					
XVII.	<i>bekommen</i> + AdjP + NP <sub>AKK</sub>	resultativ + EXERT	+	+	+	(+)	(+)
XVIII.		inchoativ					
XIX.	<i>bekommen</i> + zu(m)-Inf. (+ NP <sub>AKK</sub> )	modal	+	+	+	+	+
XX.		resultativ + EXERT	(+)				
XXI.	<i>bekommen</i> + Part. II (+ NP <sub>AKK</sub> )	passiv	+	+	+	+	+

Abb. 7: Areale Verteilung von *bekommen*-Belegen in den Zeitungskorpora, sortiert nach syntaktischen Mustern (nach Lenz 2013: Tab. 23)

2015a und 2015b). Beide Verben können im Gegenwartsdeutschen innerhalb adverbialer, adjektivischer und sogar partizipialer Resultativkonstruktionen (s. Abb. 5) auftreten (z. B. adverbial: den Schlüssel (nicht) ins Schloss kriegen/bekommen ,es (nicht) schaffen, den Schlüssel ins Schloss zu stecken'; adjektivisch: *die Kinder (nicht) satt kriegen/bekommen* ,es (nicht) schaffen, die Kinder zu sättigen'; partizipial: *die Aufgabe (nicht) gelöst kriegen/bekommen* ,es (nicht) schaffen, die Aufgabe zu lösen'). Hinsichtlich der Frequenzverteilungen der beiden Verben innerhalb der Resultativkonstruktionen zeigt sich aber, dass es selbst in standard-schriftsprachlichen Daten häufiger *kriegen* ist, das die Resultativkonstruktionen besetzt, während *bekommen* relativ gesehen deutlich seltener als resultatives Verb auftritt, und wenn es dies tut, dann fast ausschließlich in adverbialen Resultativkonstruktionen, während – die insgesamt auch selteneren – adjektivischen und partizipialen Resultativkonstruktionen vor allem *kriegen* vorbehalten sind (s. die *kriegen*-Typen XXIII, XVII und XXX in Abb. 6 bzw. die *bekommen*-Typen XIV, XVII und XX in Abb. 7).

## 5. Plädoyer für eine variationssensitive Lexikographie

Der Diskussion in diesem Beitrag wurden einleitend drei Forschungsfragen vorangestellt, die mit Fokus auf standardsprachorientierte Nachschlagewerke des Gegenwartsdeutschen beleuchtet wurden:

1. *Wie viel Variation braucht ein Wörterbuch?*
2. *Wie viel Variation verträgt ein Wörterbuch?*
3. *Wie stellt ein Wörterbuch Variation „optimal“ dar?*

Es wurde deutlich gemacht (s. Kap. 2), dass zur Beantwortung dieser Fragen zunächst zu klären ist, welches standardsprachliche Konzept einem Nachschlagewerk zugrunde gelegt wird und welche Rolle Variation in diesem Standardsprachkonzept zukommt. Dieses Standardsprachkonzept wie auch sein Verhältnis zur „Sprachrealität“, das heißt zum Standardsprach-*Gebrauch* in all seinen Facetten, sollte im Wörterbuch explizit offengelegt werden: Wie wird Standardsprache (aus Norm- und Gebrauchsperspektive) definiert und lexikographisch operationalisiert? Welche linguistischen wie sozio-pragmatischen Variationsdimensionen werden wie dokumentiert? Welche areal-horizontalen, vertikal-sozialen, medial-konzeptionellen und anderen Steuerungsparameter finden in welcher Art und Weise Berücksichtigung? Das inkludiert auch die Frage: Welche Räume und Länder werden (bewusst nicht) wie berücksichtigt (Stichwort „Plurizentrik“)? Die Entscheidung, bestimmte Variationsdimensionen in einem Nachschlagewerk zu berücksichtigen, impliziert auch die Notwendigkeit der theoretisch-konzeptionellen wie operationellen Fassung dieser Variationsdimensionen in Form terminologischer Standards, die im Wörterbuch ihren Niederschlag in einem einheitlichen, explizit offengelegten und in seiner Anwendung nachvollziehbaren Metavokabular finden. In dem Maße, in dem ein Nachschlagewerk einen deskriptiven und weniger präskriptiven Ansatz verfolgt, bedarf es der Analyse variationssensitiver Korpora, die überhaupt eine empirische Herangehensweise an die zu berücksichtigenden Variationsdimensionen erlauben. Exemplarische Analysen auf Basis eines solchen variationssensitiven Korpus wurden im Vorangehenden am Beispiel der Verben *kriegen* und *bekommen* vorgestellt (s. Kap. 4). Nicht selten sind umfangreiche korpuslinguistische Auswertungen natürlich nur schwer in die lexikographische Forschungspraxis zu integrieren, weswegen eine Verbindung der lexikographischen Arbeit im engeren Sinne mit aktueller (variationslinguistisch orientierter) Forschung anzustreben ist. Im Hinblick auf die linguistischen Steuerungsparameter, die das Auftreten eines Lemmas bedingen, sollte in jedem

Fall auch seine Einbettung in seine syntakto-semantischen Kontexte im Lexikon systematisch Berücksichtigung finden. Das DWDS-Tool „Wortprofile“ leistet in dieser Hinsicht bereits einen wertvollen Beitrag. Wie eine solche (konstruktionelle) Einbettung auch aus sprachdynamischer Perspektive visualisiert werden könnte, illustriert abschließend Abb. 8 am Beispiel von *kriegen* auf Basis sämtlicher seiner syntakto-semantischen Typen, die in den historischen wie gegenwartssprachlichen Korpora aus Lenz (2013) eruiert werden konnten.

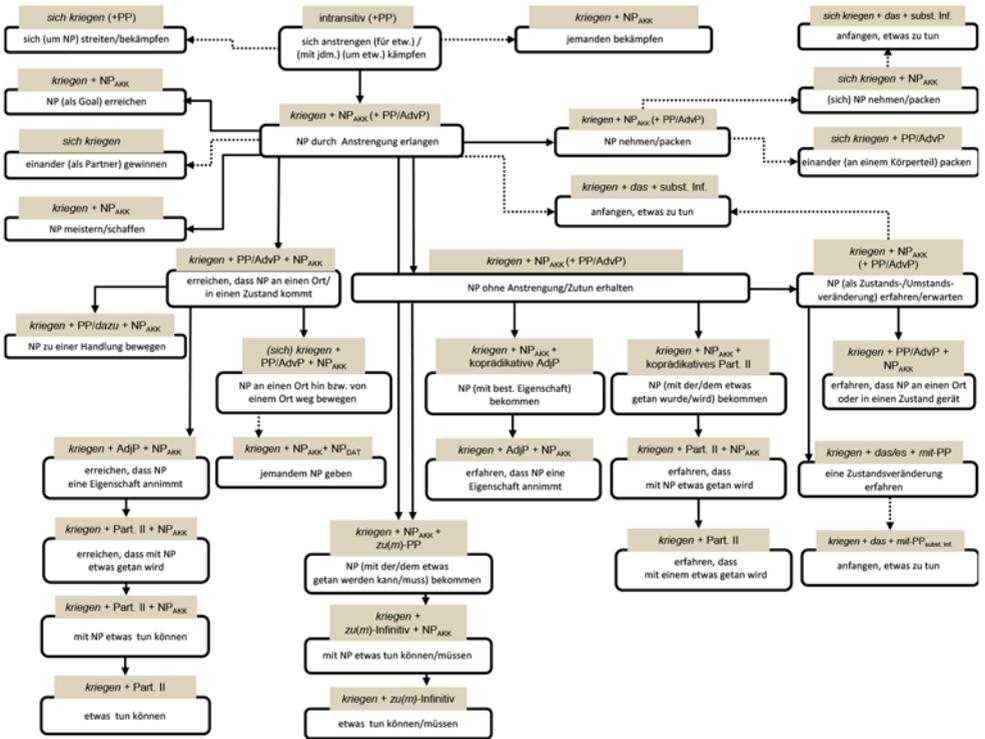


Abb. 8: Entwicklungspfade zentraler *kriegen*-Varianten (Lenz 2013: Abb. 50; gestrichelte Pfeillinien deuten auf empirisch nur schwach abgesicherte Verbindungen hin)

## Literatur

AdA = Elspaß, Stephan & Robert Möller. 2003 ff. *Atlas zur deutschen Alltagssprache*. URL: [www.atlas-alltagssprache.de](http://www.atlas-alltagssprache.de) [letzter Zugriff: 10.10.2019].

Ammon, Ulrich. 1995. *Die deutsche Sprache in Deutschland, Österreich und der Schweiz. Das Problem der nationalen Varietäten*. Berlin & New York: de Gruyter.

Ammon, Ulrich, Hans Bickel, Jakob Ebner, u.a. (Hrsg.). 2004. *Variantenwörterbuch des Deutschen. Die Standardsprache in Österreich, der Schweiz und Deutschland sowie in Liechtenstein, Luxemburg, Ostbelgien und Südtirol*. Berlin & Boston: de Gruyter.

Ammon, Ulrich, Hans Bickel & Alexandra N. Lenz (Hrsg.). 2016. *Variantenwörterbuch des Deutschen. Die Standardsprache in Österreich, der Schweiz, Deutschland, Liechtenstein, Luxemburg, Ostbelgien und Südtirol sowie Rumänien, Namibia und Mennonitensiedlungen*. 2., völlig neu bearbeitete und erweiterte Auflage. Berlin & Boston: de Gruyter.

Auer, Peter. 2018. The German neo-standard in a Europa context“. In Gerhard Stickel (Hrsg.), *National Language Institutions and National Languages*, 37–56, Budapest: Hungarian Academy of Science/European Federation of National Institutions for Language.

Barbour, Stephen & Patrick Stevenson. 1998. *Variation im Deutschen. Soziolinguistische Perspektiven*. Berlin & Boston: de Gruyter.

Bickel, Hans & Christoph Landolt. 2018. *Schweizerhochdeutsch. Wörterbuch der Standardsprache in der deutschen Schweiz*. 2. Auflage. Mannheim: Dudenverlag.

Clyne, Michael G. 1989. Pluricentricity. National Variety. In Ulrich Ammon (Hrsg.), *Status and Function of Languages and Language Varieties*, 357–371. Berlin & New York: de Gruyter.

Datenbank Gesprochenes Deutsch (DGD) Herausgegeben vom Leibniz-Institut für Deutsche Sprache (IDS). URL: <https://dgd.ids-mannheim.de/> [letzter Zugriff: 09.10.2019].

Deppermann, Arnulf, Stefan Kleiner & Ralf Knöbl. 2013. 'Standard usage': Towards a realistic conception of spoken standard German. In Peter Auer, Javier Caro Reina & Göz Kaufmann (Hrsg.), *Language Variation – European Perspectives IV. Selected papers from the Sixth International Conference on Language Variation in Europe (IC-LaVE 6)*, Freiburg, June 2011, 83–116. Amsterdam/Philadelphia: Benjamins (SILV 14).

Duden. 2015. *Deutsches Universalwörterbuch*. 8. überarbeitete und erweiterte Auflage. Berlin: Dudenverlag.

Duden. 2016. *Das Wörterbuch der sprachlichen Zweifelsfälle: Richtiges und gutes Deutsch*. Berlin: Bibliographisches Institut.

DWDS – Digitales Wörterbuch der deutschen Sprache. Das Wortauskunftssystem zur deutschen Sprache in Geschichte und Gegenwart, hrsg. v. d. Berlin-Brandenburgischen Akademie der Wissenschaften. URL: <https://www.dwds.de/> [letzter Zugriff: 10.10.2019].

Ebner, Jakob. 2009. Duden – Wie sagt man in Österreich? Wörterbuch des österreichischen Deutsch. 4., völlig überarb. Aufl. Wien et al.: Dudenverlag.

Eichinger, Ludwig & Werner Kallmeyer (Hrsg.). 2005. *Standardvariation: Wie viel Variation verträgt die deutsche Sprache?* Berlin & Boston: de Gruyter.

WDG = Klappenbach, Ruth & Wolfgang Steinitz (Hrsg.). 1967. Wörterbuch der deutschen Gegenwartssprache. 1. Band: A-deutsch. Dritte, durchgesehene Auflage. Berlin: Akademie Verlag.

Glauning, Manfred Michael. 2013. Deutsch im 21. Jahrhundert: „pluri“-, „supra“- oder „postnational“? In Doris Sava & Hermann Scheuringer (Hrsg.), *Dienst am Wort: Festschrift für Ioan Lazarescu zum 60. Geburtstag* (Forschungen zur deutschen Sprache in Mittel-, Ost- und Südosteuropa 3), 123–132. Passau: Stutz.

Herrgen, Joachim. 2015. Entnationalisierung des Standards: Eine perzeptionslinguistische Untersuchung zur deutschen Standardsprache in Deutschland, Österreich und der Schweiz. In Alexandra N. Lenz & Manfred Michael Glauning (Hrsg.), *Standarddeutsch im 21. Jahrhundert – Theoretische und empirische Ansätze mit einem Fokus auf Österreich* (Wiener Arbeiten zur Linguistik 1), 139–164. Göttingen: Vienna University Press.

Klein Wolf P. 2013. Warum brauchen wir einen klaren Begriff von Standardsprachlichkeit und wie könnte er gefasst werden? In Jörg Hagemann, Wolf P. Klein & Sven Staffeldt (Hrsg.), *Pragmatischer Standard*, 15–33. Tübingen: Staffenburg.

Kleiner, Stefan. 2014. Die Kodifizierung der deutschen Standardausprache im Spiegel der faktischen Variabilität des Gebrauchsstandards. In Albrecht Plewnia & Andreas Witt (Hrsg.), *Sprachverfall? Dynamik – Wandel – Variation* (Jahrbuch des Instituts für Deutsche Sprache 2013), 273–298. Berlin & Boston: de Gruyter.

Koppensteiner, Wolfgang & Alexandra N. Lenz. 2017. Theoretische und methodische Herausforderungen einer perzeptiv-attitudinalen Standardsprachforschung: Perspektiven aus und auf Österreich. In Heinz Sieburg & Hans-Werner Solms (Hrsg.), *Das Deutsche als plurizentrische Sprache: Ansprüche – Ergebnisse – Perspektiven* (Zeitschrift für deutsche Philologie. Sonderheft 136), 43–68. Berlin: Erich Schmidt Verlag.

Koppensteiner, Wolfgang & Alexandra N. Lenz [eingereicht]. *Tracing a Standard Language in Austria Using Methodological Microvariations of Verbal and Matched Guise Technique*.

Janwermeyer, Manuela, Johanna Fanta-Jende, Alexandra N. Lenz & Katharina Kocrecky-Kröll. 2019. Competing norms of standard pronunciation. Phonetic analyses of the <-ig>-variation in Austria. [to appear in *Dialectologia et Geolinguistica*]

Lenz, Alexandra N. 2013. Vom >kriegen< und >bekommen<. Kognitiv-semantische, variationslinguistische und sprachgeschichtliche Perspektiven. (Linguistik. Impulse&Tendenzen 53) Berlin & New York: de Gruyter.

Lenz, Alexandra N. 2014. Sprachvariation und Sprachwandel aus der Perspektive von Deutschlehrerinnen und Deutschlehrern – Einstellungsdaten aus Österreich, Deutschland und der Schweiz. In Albrecht Plewnia & Andreas Witt (Hrsg.), *Sprachverfall? Dynamik – Wandel – Variation*, (Jahrbuch des Instituts für deutsche Sprache 2013), 323–371. Berlin: de Gruyter

Lenz, Alexandra N. 2015a. Niederdeutsche Resultativ-, Modal- und Kausativkonstruktionen. Panchronische Analysen. *Jahrbuch des Vereins für Niederdeutsche Sprachforschung* 138. 77–97.

Lenz, Alexandra N. 2015b. On the resultative-modal grammaticalisation pathway of German GET verbs – with an outlook on Dutch and Afrikaans. *Taal en Tongval* 67(2). 177–209.

Niebaum, Hermann & Jürgen Macha. 2006. *Einführung in die Dialektologie des Deutschen*. Berlin & Boston: de Gruyter.

ÖWB = Österreichisches Wörterbuch. 2018. Herausgegeben im Auftrag des Bundesministeriums für Bildung. 43. Auflage. Wien: öbv.

Reichmann, Oskar [u. a.]. 1988. Zur Vertikalisierung des Varietätenspektrums in der jüngeren Sprachgeschichte des Deutschen. In Horst Haider Munske, Peter von Polenz, Oskar Reichmann & Reiner Hildebrandt (Hrsg.), *Deutscher Wortschatz. Lexikologische Studien. Ludwig Erich Schmitt zum 80. Geburtstag von seinen Marburger Schülern*, 151–180. Berlin & New York: de Gruyter.

Scharloth, Joachim. 2005. *Sprachnormen und Mentalitäten. Sprachbewusstseinsgeschichte in Deutschland im Zeitraum von 1766 und 1785* (Reihe Germanistische Linguistik 255). Tübingen: Niemeyer.

Scheuringer, Hermann. 1996. Das Deutsche als pluriareale Sprache: Ein Beitrag gegen staatlich begrenzte Horizonte in der Diskussion um die deutsche Sprache in Österreich. *Die Unterrichtspraxis / Teaching German* 29(2). 147–153.

Schmidlin, Regula. 2011. *Die Vielfalt des Deutschen: Standard und Variation. Gebrauch, Einschätzung und Kodifizierung einer plurizentrischen Sprache* (Studia Linguistica Germanica 106). Berlin & Boston: de Gruyter.

Wiesinger, Peter. 2015. Das österreichische Deutsch in der globalisierten Umwelt: Wandlungen durch bundesdeutsche Einflüsse. In Alexandra N. Lenz, Timo Ahlers & Manfred M. Glauning (Hrsg.), *Dimensionen des Deutschen in Österreich. Variation und Varietäten im sozialen Kontext* (Schriften zur deutschen Sprache in Österreich 42), 91–122. Frankfurt/Main: Peter Lang.

### Zeitungstexte

(abgefragt über das Recherche- und Analysesystem Cosmas IIweb (Version 1.6.2) des Instituts für deutsche Sprache (Mannheim) [URL: <<http://www.ids-mannheim.de/cosmas2/>>])

HannAllg = Hannoversche Allgemeine [URL: <<http://www.haz.de/>>]  
NieÖ = Niederösterreichische Nachrichten [URL: <<http://www.noen.at/>>]  
NÜN = Nürnberger Nachrichten [URL: <<http://www.nordbayern.de/nuernberger-nachrichten/>>]  
RhZ = Rhein-Zeitung [URL: <<http://www.rhein-zeitung.de/startseite.html>>]

**Lars Trap-Jensen** ist leitender Redakteur bei ordnet.dk, verfügt über mehr als 25 Jahre Erfahrung in der praktischen Lexikographie und betreut die digitale Ausgabe des Dänischen Wörterbuchs und des Wörterbuchs der dänischen Sprache sowie die Belegstellen-Ressource KorpusDK. Er war und ist Mitglied in verschiedenen lexikographischen Verbänden im In- und Ausland.



Foto: BBAW, Judith Affolter

## **SPRACHPOLITIK IST KULTURPOLITIK: GEGENWARTSLEXIKOGRAPHIE ALS ÖFFENTLICHE DIENSTLEISTUNG IN SKANDINAVIEN**

Die lexikographischen Traditionen in Deutschland und in den skandinavischen Ländern haben mehr gemeinsam, als viele im Allgemeinen annehmen. Die germanischen Sprachen bilden nicht nur einen gemeinsamen Sprachraum, sondern haben auch kulturell und wissenschaftsgeschichtlich sehr viele Ähnlichkeiten. Auf dem Gebiet der Lexikographie lässt sich eine relativ parallele Entwicklung der nationalen Wörterbuchwerke in den germanischsprachigen Ländern verfolgen.

So findet man unter den historischen Wörterbüchern in England das *Oxford English Dictionary*, in den Niederlanden das *Woordenboek van der nederlandsche Taal* und in Deutschland Jacob und Wilhelm Grimms *Deutsches Wörterbuch*. Diese Werke sind allgemein bekannt, aber welche Parallelen gibt es in den skandinavischen Ländern?

In Schweden wurde die Vorbereitung eines nationalen Wörterbuchs schon bei der Gründung der schwedischen Akademie im Jahr 1786 in ihrer Satzung festgelegt. So heißt es in Ziffer 22, dass „die wichtigste Aufgabe der Akademie“ darin bestehe,

„sich für die Reinheit, Stärke und Hoheit der schwedischen Sprache einzusetzen“. Heute würde man wohl eher sagen: für die Klarheit, die Ausdruckskraft und das Ansehen. Wie dem auch sei, die schwedische Akademie arbeitet seitdem an dem Wörterbuch, die ersten 100 Jahre zwar nur mit relativ mäßigem Engagement, der erste Band erschien 1898, aber danach sind die Bände in regelmäßigem Rhythmus herausgebracht worden. Bisher sind 36 Bände und zwei Lieferungen des Bands 37 bis zum Wort „vret“ („Grundstück, Bodenparzelle“) veröffentlicht, und die Vollendung des Werkes wird voraussichtlich 2024 erfolgen.

In Dänemark und Norwegen begann die Entwicklung etwas später und mit weniger ehrgeizigen Projekten, die sich jedoch eindeutig von der Entwicklung in den eben genannten Ländern inspirieren ließen: In Dänemark wurde das *Ordbog over det danske Sprog* (Wörterbuch der dänischen Sprache) in der Zeit von 1918 bis 1956 herausgegeben und erschien zunächst in 28 Bänden, zu denen sich später fünf Ergänzungsbände hinzufügten. Insgesamt umfasste es etwa 220.000 Stichwörter. In Norwegen wurden in den 1930er Jahren gleich zwei nationale Wörterbuchprojekte begonnen, da es in Norwegen zwei Hauptvarianten der Schriftsprache gibt. So wurde in den Jahren 1937–1957 das norwegische *Riksmål-Wörterbuch* veröffentlicht, zu dem später noch zwei Ergänzungsbände erschienen, während die Arbeit an dem Wörterbuch für die Variante Nynorsk, *Norsk Ordbok* (Norwegisches Wörterbuch) 1930 begann und erst 2016 abgeschlossen wurde. Insgesamt umfasst letztgenanntes Werk 300.000 Stichwörter in zwölf Bänden.

Auch im Rahmen der sogenannten digitalen Revolution während der vergangenen zwei Jahrzehnte weisen die nationalen Wörterbuchprojekte gemeinsame Merkmale auf. Während dieser ‚zweiten Welle‘ wurden die ursprünglich gedruckten Wörterbücher in digitale Formate umgesetzt, retrodigitalisiert und es wurden in mehreren Ländern zum Teil auch neue Projekte ins Leben gerufen. Dahinter steht die Erkenntnis, dass die großen alten Wörterbuchwerke aus synchroner Perspektive veraltet waren. Dies führte zu neuen Projekten in den Niederlanden, in Großbritannien, Schweden, Dänemark und teilweise in Norwegen und Deutschland. Beispiele sind das *Algemeen Nederlands Woordenboek* in den Niederlanden, OxfordDictionaries.com (eine Sammlung von Wörterbüchern, die inhaltlich u. a. auf dem *Oxford Dictionary of English* und dem *New Oxford American Dictionary* basiert) in Großbritannien, und das DWDS in Deutschland – all dies sind Projekte, die im Rahmen des ZDL natürlich bekannt sind. Weniger bekannt sind vielleicht einige skandinavische Projekte, von denen ich deswegen die prominentesten kurz vorstellen möchte.

## Skandinavische Gegenwartslexikographie

In Schweden wurde ein modernes schwedisches Wörterbuch in Verbindung mit der schwedischen Nationalenzyklopädie erstellt. Es wurde in drei Bänden von 1995 bis 1996 veröffentlicht und ist später unter dem Titel *Svensk ordbok* („Schwedisches Wörterbuch“) erschienen, von Redakteuren der lexikographischen Abteilung der Universität Göteborg erarbeitet und von der Schwedischen Akademie finanziert.

Abbildung 1 zeigt die digitalen Ausgaben der drei Wörterbücher der Akademie. Links die SAOL (Abkürzung von *Svenska Akademiens ordlista*, „Wortliste der schwedischen Akademie“), ein orthographisches Wörterbuch mit etwa 120.000 Stichwörtern und deren offiziellen Schreibweisen. In der Mitte das eben erwähnte *Svensk ordbok* in digitaler Form und rechts das historische Wörterbuch der schwedischen Akademie. Diese drei Wörterbücher werden den Benutzern kostenlos zur Verfügung gestellt. Das *Svensk ordbok* wird regelmäßig, aber in begrenztem Umfang, aktualisiert. Der Betrieb der Website und die Entwicklung des Wörterbuchs sind bis 2060 durch eine Vereinbarung zwischen der Akademie und der Universität Göteborg gewährleistet.



Abb. 1: Portalseite der drei schwedischen Akademiewörterbücher, svenska.se

In Norwegen gibt es zwei erwähnenswerte moderne Wörterbuchprojekte. Zunächst das *Wörterbuch der norwegischen Akademie NAOB*, das 2018 erschienen ist. Dies ist eine digitale und völlig modernisierte Version des eingangs erwähnten historischen *Riksmål-Wörterbuchs*, das ursprünglich seit den 1930er Jahren erarbeitet und in einem etwas unregelmäßigen Rhythmus in Einzelteilen herausgegeben wurde (siehe Abbildung 2).



**Abb. 2:** Startseite des Wörterbuchs der norwegischen Akademie, *naob.no*

Das NAOB-Wörterbuch ist also in erster Linie ein Digitalisierungsprojekt, bei dem das Originalwerk mit seinen Ergänzungsbänden retrodigitalisiert und zu einem digitalen Wörterbuch zusammengesetzt wurde. Darüber hinaus wurde die Artikeldatei mittels einer sowohl strukturellen wie inhaltlichen Überarbeitung und Ergänzung modernisiert. Das Ergebnis ist ein Wörterbuch mit etwa 225.000 Stichwörtern in aktualisierter, Bokmål-naher Orthographie.

Neben dieser Arbeit gibt es auch ein kleineres Wörterbuch, das die beiden offiziellen Standardsprachen Bokmål und Nynorsk abdeckt und ursprünglich als Gemeinschaftsprojekt des Norwegischen Sprachrats und der Universität Oslo entstand. Damals, im Jahre 1986, handelte es sich um zwei eigenständige gedruckte Wörterbücher, die jedoch mit identischer Struktur und gleichem Inhalt für die beiden Hauptvarianten des Norwegischen parallel erarbeitet wurden. Zwischenzeitlich haben sie sich aber etwas unterschiedlich entwickelt, was insbesondere aus dem Stichwortbestand hervorgeht: So gibt es für Bokmål 70.000 Stichwörter, aber 90.000 für Nynorsk. Der Unterschied macht sich vor allem im Moment der digitalen Nutzung bemerkbar, bei der man die Artikel aus beiden Werken parallel betrachten kann. Eine Überprüfung der beiden Wörterbücher wurde im vergangenen Jahr eingeleitet, um den Wortbestand beider Wörterbücher zu vereinheitlichen und auf 100.000 zu erhöhen. Das Projekt ist auf fünfzehn Jahre angelegt und wird an der Universität Bergen durchgeführt, die im Zusammenhang mit einer umfassenden Neuorganisation der lexikographischen Arbeit in Norwegen die redaktionelle Verantwortung dafür übernommen hat (siehe Abbildung 3).

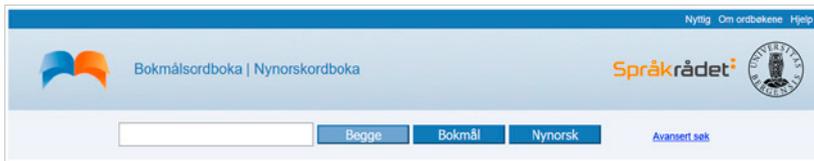


Abb. 3: Startseite des *Bokmåls-/Nynorskordboka*, <https://ordbok.uib.no>

In Dänemark deckt die Website *ordnet.dk*, und hier insbesondere *Den Danske Ordbog*, DDO (Das Dänische Wörterbuch), den modernen Wortschatz ab. DDO wurde ursprünglich als gedrucktes Wörterbuch in sechs Bänden 2002–2005 veröffentlicht, wurde jedoch bereits bei Aufnahme der redaktionellen Arbeiten im Jahre 1991 für die digitale Weiterverwendung konzipiert und als komplex strukturiertes SGML-/XML-Dokument erstellt. Es war außerdem das erste korpusbasierte Wörterbuch des Dänischen.

Seit 2009 ist eine digitale Version im Web unter *ordnet.dk* verfügbar, wo Nutzer ebenfalls auf eine retrodigitalisierte Version des historischen *Ordbog over det danske Sprog* sowie auf Korpusmaterial zugreifen können. Da es ein ständig aktualisiertes Wörterbuch mit der Möglichkeit, auf Korpora zuzugreifen, kombiniert, ist das *ordnet.dk*-Projekt dem heutigen DWDS-Projekt sehr ähnlich (siehe Abbildung 4).

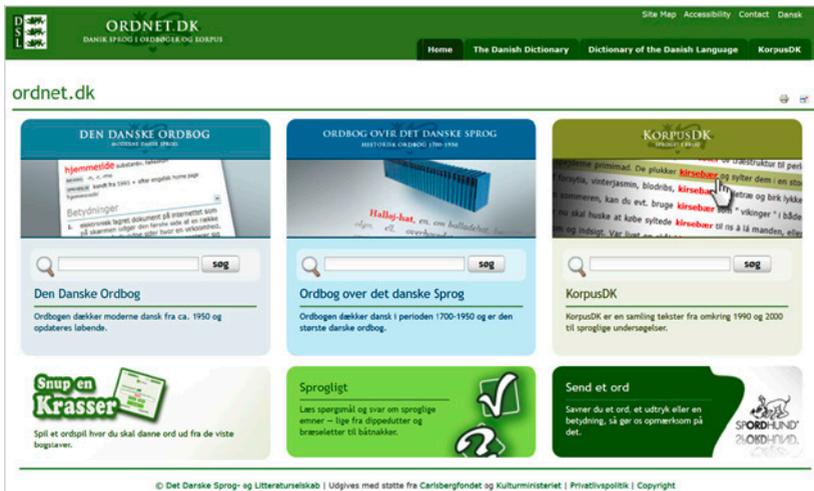


Abb. 4: Startseite des *ordnet.dk*

Die großen historischen Nationalwörterbücher dürften kaum jemals wirtschaftlich profitabel gewesen sein. So wird berichtet, dass das *Oxford English Dictionary*, obwohl es von Hunderttausenden, wenn nicht Millionen genutzt wird, noch nie einen Gewinn für den Verlag Oxford University Press erzielt hat (so Judy Pearsall in einem persönlichen Gespräch). Es versteht sich fast von selbst: Je kleiner ein Markt ist, desto schwieriger ist es, Gewinne zu erzielen. In Skandinavien ist es auch immer so gewesen, dass Wörterbücher mit über etwa 60.000 Stichwörtern nicht rentabel waren; sämtliche der hier vorgestellten Werke wurden deshalb aus öffentlichen Mitteln finanziert, entweder direkt aus der Staatskasse oder aus Stiftungen mit kulturellem oder wissenschaftlichem Auftrag. In diesem Sinne ist die Dokumentation und Beschreibung des Wortschatzes nach wie vor eine kulturpolitische Aufgabe und eine öffentliche Verantwortung.

Ich möchte in diesem Zusammenhang auf ein letztes Projekt hinweisen, das eindeutig zeigt, dass Wörterbucharbeit Kulturpolitik ist. Die färöische Website *sprotin.fo* (vgl. Abbildung 5), die 2018 online ging, wurde öffentlich finanziert, jedoch in Zusammenarbeit mit einem Verlag, der für den Betrieb der Website verantwortlich ist. Hier haben färöische Benutzer Zugriff auf eine ganze Reihe ein- und zweisprachiger Wörterbücher, darunter sehr gute Wörterbücher, denn die Färinger waren sich immer der Rolle der Sprache als nationales Symbol bewusst. So ist z. B. das Deutsch-Färöische Wörterbuch ein außerordentlich umfangreiches Werk mit fast 150.000 Stichwörtern. Es ist beeindruckend und bewundernswert, dass so etwas in einem kleinen Land mit 50.000 Einwohnern, die gleichzeitig auch fast alle färöischen Sprecher ausmachen, überhaupt möglich ist.



Abb. 5: Startseite des färöischen Wörterbuchportals *sprotin.fo*

Aber zurück zu den großen Nationalwörterbüchern und ihrer Rolle als Kulturträger: Eine der gewaltigsten Veränderungen durch die Digitalisierung, die nicht nur die erwähnten germanischsprachigen Wörterbücher betrifft, sondern

ein allgemeiner Trend in der Vermittlung kultureller Inhalte zu sein scheint, ist die, dass es kommerziellen Verlegern kaum noch möglich ist, mit der Produktion von Wörterbüchern Gewinne zu erzielen. Überall in Europa ist zu sehen, wie traditionelle Wörterbuchverlage ihre Marktposition nicht oder nur unter großen Schwierigkeiten behaupten können; viele müssen ihre Wörterbuchsparte entweder ganz aufgeben oder erheblich reduzieren. Auffällige europäische Beispiele sind *Chambers Harrap* in Großbritannien und *Duden* sowie *Wahrig* in Deutschland; und in Dänemark hatte der Verlag *Politiken* zunächst die Wörterbuchsparte stillgelegt und sie 2018 verkauft. Dies hat allmählich zu der Erkenntnis geführt, dass die öffentliche Hand eine größere Verantwortung übernehmen muss, wenn weiterhin aktuelle und qualitativ hochwertige Wörterbücher verfügbar sein sollen. Für die Wörterbucharbeit der Zukunft sollte dabei auch berücksichtigt werden, dass ein großer Bedarf an strukturierten und formalisierten Sprachdaten für sprachtechnologische Zwecke besteht. Denn Spracherkennung, Sprachsynthese, maschinelle Übersetzung usw. sind nur begrenzt möglich ohne geeignete lexikalisch-semantische Daten. Bei der Erarbeitung der Datengrundlage der Wörterbücher der Zukunft sollte auch dem sprachtechnologischen Aspekt Rechnung getragen werden.

Ich möchte deshalb abschließend noch einmal betonen, wie wichtig es gerade jetzt ist, dass Lexikographen, Computerlinguisten und Informatiker zusammenarbeiten, um lexikographische Daten in Standardformaten zusammenzustellen und diese untereinander und mit Kollegen außerhalb zu teilen. Auch für diese Aufgaben dürfte das ZDL hervorragend geeignet sein, da es schon im Grundsatz als eine Zusammenarbeit zwischen verschiedenen Arbeitsstellen und Mitarbeitern aus unterschiedlichen Disziplinen konzipiert ist.

Ich möchte zuletzt noch auf die EU-Infrastruktur ELEXIS aufmerksam machen. ELEXIS zielt darauf ab, solche Bemühungen zu harmonisieren, Werkzeuge zu entwickeln, die von allen genutzt werden können und dadurch die Kosten und den Zeitaufwand für die Aktualisierung vorhandener oder die Entwicklung neuer Ressourcen zu reduzieren. In der ELEXIS-Initiative werden gemeinsame Standards eingeführt, um inkompatible Daten kompatibel zu machen, mit der Perspektive, vorhandene Ressourcen miteinander zu verknüpfen.

Ich wünsche allen Mitarbeitern und anderen Beteiligten viel Erfolg bei der bevorstehenden Arbeit. Wir vom Projekt *ordnet.dk* werden die Arbeit des ZDL mit großem Interesse verfolgen und ich hoffe, dass unsere beiden Wörterbuchredaktionen auch künftig Erfahrungen, Herausforderungen und Ergebnisse unserer Arbeit miteinander teilen und besprechen können.

**Sarah Ogilvy** ist Senior Research Fellow an der Fakultät für Linguistik, Philologie und Phonetik sowie Direktorin des Dictionary lab an der Universität Oxford. Zuvor arbeitete sie im Innovation Lab bei Amazon, später als Editorin verschiedener großer Wörterbücher, unter anderem des Oxford English Dictionary, und lehrte in Cambridge und Stanford. Als Linguistin, Lexikographin und Informatikerin interessiert sie sich besonders für die Schnittstelle von Technologie und Geisteswissenschaften.



Foto: BBAW, Judith Affolter

## **DIGITAL LEXICOGRAPHY AND OXFORD DICTIONARIES: THE INTERNATIONAL CONTEXT**

It is a great honour to be here today to celebrate the launch of the Centre for Digital Lexicography of the German Language, and to carry messages of congratulations and support from your colleagues across the channel at Oxford Dictionaries in England. In particular, I thank Martin Grötschel and Andreas Gardt (presidents of the Academies in Berlin and Göttingen respectively); Wolfgang Klein, Alexander Geyken, Volker Harm, and your respective teams. It is excellent that your funders, the Federal Ministry of Education and Research, appreciate the value of digital language resources, and are supporting and investing in the centre. By doing so, they are investing in the future and we hope that other governments around the world might be inspired to follow their initiative.

This evening, I will share with you some reflections on the influence of German lexicography on the history of my own tradition of English lexicography, and then

speak a little about some of the digital collaborations that Oxford University Press is forging around the globe with academics, institutions, industry, and language communities, and our hope for future collaborations between Oxford and you here in Berlin.

There has always been cultural and linguistic exchange between Germany and Britain. The nineteenth century was a particularly fertile period with the English language borrowing many words from German, e.g. semester" (1826), ozone" (1841), and aspirin" (1899). The exchange was mutual with German also borrowing English words such as toast" (1803), reporter" (1838), and schrapnell" (1868). Many English speakers may think that the word shrapnel" sounds German and therefore came from German, but in fact the Shrapnel shell was invented by a British army officer in the 1790s called Henry Shrapnel (1761–1842).

However, I would like to offer more than an exchange of lexicon; this evening I would like to remember other forms of exchange between our two nations, in particular that of lexicographic methodologies, policies, and practices.

In 1700 Gottfried Leibniz declared that language must play a central role in the research focus of the then newly-founded Prussian Academy of Sciences, and the Berlin-Brandenburg Academy of Sciences and Humanities which hosts us today has been admirably true to that mission. The Academy has hosted ambitious projects such as the Dictionary of the Brothers Grimm (*Deutsches Wörterbuch*, DWB), the *Thesaurus Latinarium* (1890s), the Dictionary of Contemporary German (*Wörterbuch der deutschen Gegenwartssprache* WDG), the Etymological Dictionary (*Etymologisches Wörterbuch*), and now the Digital Dictionary of the German Language (*Digitales Wörterbuch der deutschen Sprache* DWDS) which comprises all these dictionaries plus corpora and computational analysis.

Over those 300 years all of these impressive lexicographic projects have been watched with admiration by your friends across the channel. Indeed, Britain continues to be inspired by what you are doing here. So too were our predecessors in the nineteenth century at the London Philological Society. It was three members of the Philological Society who in the 1850s proposed the creation of the *New English Dictionary* which became known as the *Oxford English Dictionary*. It was the first English dictionary to do three things simultaneously: it was historical rather than synchronic; it was descriptive rather than prescriptive; and it was crowd-sourced by the public rather than the work of a sole editor.

It may have been the first to do this in the English language but it was certainly not the first in the world. That honour belongs to Germany.

This was brought home to me two years ago, when I visited the superb Grimmwelt museum in Kassel. This is one of my favourite museums, and if you have not visited it, you are missing out on something very special. The museum tells the story of the lives and work of Jacob Grimm (1785–1863) and his brother Wilhelm (1786–1859) (the Brothers Grimm).

Walking through the museum I immediately recognized the similarities between the OED and the Dictionary of the Brothers Grimm. I was struck by the ways in which the OED editors, who began work on their dictionary 20 years after the Grimms, had borrowed the working methods of the Brothers Grimm. I wondered whether the OED would have existed without the Grimms' dictionary, and this is the subject of research that I am currently undertaking. My research on this topic has benefited greatly from my conversations with your own Volker Harm and one of his previous staff, Wiebke Blanck.

It is precisely this kind of scholarly exchange, collaboration, and partnering that I want to celebrate this evening, as I reflect on the evolution of lexicography from the nineteenth century until the present day, and the ways in which the uses of dictionaries are changing due to technological innovation and the new lexicographic policies and practices that they prompt.

Lexicographers are finding themselves having to respond to changing uses of dictionary data: a move from the static dictionary text to new digital needs, the development of apps and APIs, and now increasingly the support of machine learning and artificial intelligence.

The digital version of the OED now has the *Historical Thesaurus* of the OED integrated within it so that users can not only read a definition of a word and learn about the life of that word across time, but also link through to the historical thesaurus which tells you about that word's semantic domain and its synonyms across time.

Let me share with you some ways in which our digital data is bringing us into collaborations and partnerships with others around the world, be that with academics, with institutions and industry, and directly with language communities.

As I talk about these collaborations I hope it might spark for you ideas about how you and your new centre might want to collaborate with us in Oxford.

### **Collaborations with Academics**

At the OED, we are keen to respond to scholars and support them in their research needs, and we work closely with academics to ensure that our data is useful to them and their research projects. We also share our data with researchers via an API, and we host scholars who want to visit us to exchange ideas, data, and digital tools.

We are particularly keen to explore digital tools and methods which aid us in our dictionary-editing processes. We are mindful of what Leibniz said in 1685 when describing the value to astronomers of his *machina arithmetica*, the hand-cranked calculating machine he had invented in 1673: “It is unworthy of excellent men to lose hours like slaves in the labour of calculations which could safely be relegated to anyone else if machines were used”.

To that end, the Digital Humanities Institute at the University of Sheffield is currently working with the OED in digital experimentation to prototype a quotation retrieval interface that will enable dictionary editors to search for quotations of multi-sense words across historical corpora.

### **Collaborations with Industry**

The OED is also collaborating with institutions and industry such as this partnership called Pret a LLOD, which is funded by the European Union to work on linked linguistic open data. It is a three-year project comprising a consortium of ten companies and universities.

### **Collaborations with Language Communities**

OUP produces dictionaries in many languages other than English, and one particularly important recent initiative is the Oxford Global Languages programme which is a collaboration between Oxford Dictionaries and language communities around the world.

Focussing on one hundred languages that are currently under-resourced and under-represented on the internet, this programme works with local communities to build dictionaries and language resources which are then provided for free on the internet. The response has been very positive, and language champions in various countries have inspired hundreds of volunteers to help build the dictionaries. Begun four years ago, the programme has so far launched dictionaries and websites in 18 languages from Zulu, Malay and Urdu to Tamil, Southern Quechua, and Indonesian. Indonesian was particularly successful thanks to the great energy and dedication of Dr Deny Arnos Kwary who works at the University of Trunojoyo on the island of Madura in Indonesia. In 2016, Dr Kwary organized an event called the Bilingual Vocabulary Challenge which invited volunteers to work together to create entries for the Indonesian dictionary. In two hours, the group generated 800 entries.

## **Conclusion**

We have seen some examples of ways that Oxford's Dictionaries Division is moving away from seeing the dictionary as a discrete printed text towards seeing it as digital language data with unlimited functions and uses by forging collaborations with scholars, industry, and local communities. And on this auspicious occasion of the launch of your centre for digital lexicography of the German Language, may I extend warmest congratulations and best wishes, and an open invitation to work together in the future.

Perhaps we could let Leibniz have the final word: in 1714 in *Principles of Nature and Grace*, he wrote: "The present is big with the future and laden with the past." As historical lexicographers, we know only too well how the words in our dictionaries are indeed laden with the past. Our challenge now is to realise a future worthy of developments in technology and computational methods. Our challenge is to forge developments in those technologies through our digital lexicographic work. All of us at Oxford look forward to sharing that future with you. Thank you.

**Klaus-Dieter Lehmann** ist Professor für Bibliothekswissenschaft und Präsident des Goethe-Instituts, zudem Mitglied der Berlin-Brandenburgischen Akademie der Wissenschaften sowie der Akademie der Wissenschaften und Literatur Mainz, Vorsitzender des Stiftungsbeirats und Stiftungsratsmitglied der Kulturstiftung des Bundes und Stellvertretender Vorsitzender des Kuratoriums Kulturstiftung der Länder. Sein Engagement für die deutsche Sprache und Kultur wurde mit verschiedenen Auszeichnungen gewürdigt, unter anderem mit dem Großen Bundesverdienstkreuz und dem Verdienstorden des Landes Berlin.



Foto: BBAW, Judith Affolter

## **DIE DEUTSCHE SPRACHE, IHR WORTSCHATZ UND INTERNATIONALE SPRACHARBEIT**

Die deutsche Sprache ist nicht *die* Weltsprache – hier werden wir der englischen Sprache den Rang als Lingua franca nicht streitig machen – aber ohne Zweifel ist sie eine der bedeutendsten Kultursprachen in der Welt. Ohne das Englische als internationaler Kongress- und Publikationssprache kommen wir mittlerweile nicht mehr aus. Sie ist eine internationale Verkehrssprache geworden. Das hat sicher auch zu einer Bereicherung des internationalen Wissensaustausches geführt. Aber die Beschränkung auf eine Lingua franca bedeutet immer auch eine kognitive Einschränkung und vor allem den Ausschluss von Laien. Ich halte es bei der Sprache mit Wilhelm von Humboldt, der sinngemäß gesagt hat, jede Sprache, die ich erlerne, öffnet mir eine neue Welt.

Der Weg, den die deutsche Sprache gegangen ist, ist von einer hohen Wertschätzung der sprachlichen Vielfalt geprägt. In der langen Sprachgeschichte gab es keine zentrale Instanz, keine Akademie, keine Einzelperson, die die Regeln bestimmte. So wie etwa in Frankreich, wo die Französische Revolution die

einheitliche Sprache postulierte und sie über die Akademie kontrollierte. Die Einsicht, dass Sprache und Denken eng miteinander verbunden sind, wurde als Grundsatz der Erziehungspolitik für die Vereinheitlichung und gegen abweichende Dialektformen durchgesetzt. Damit sollte ein Bekenntnis zur Republik erreicht werden.

Deutschland war den anderen Weg gegangen, den Weg, den Martin Luther einmal so beschrieben hat: dem Volk aufs Maul schauen. Jürgen Trabant hat die sprachliche Entwicklung als Sprachliebe bezeichnet. Und in der Tat, Deutschlands bedeutendste Sprachdenker – Leibniz, Herder und Humboldt – hatten Respekt und Interesse für poetisches Denken und die Verschiedenheit der Sprachen gelehrt. Deutschland betrieb die Entwicklung der Hochsprache, ohne die Sympathie für die Vielfalt der Dialekte zu opfern.

Zwar stand am Anfang Martin Luther als kraftvoller Sprachschöpfer, aber die neuhochdeutsche Schriftsprache wurde durch verschiedene Einflüsse mitgeprägt, so durch den deutschen Südwesten mit seinen Buchdruckern, der vielfach Sprachnormen setzte, aber auch andere Entwicklungen trugen dazu bei. Deutsch ist so etwas wie das Abbild unseres Föderalismus, eine ungemein lebendige und vielfältige Sprache mit einem reichhaltigen und differenzierten Wortschatz.

Nicht zuletzt deshalb wurde die rasante Entwicklung der Wissenschaften im 19. und frühen 20. Jahrhunderts durch die Wissenschaftssprache Deutsch befördert. Sie ist ein historisch gewachsenes, traditionsreiches und komplexes Gut, präzise in den Ausdrucksformen der exakten Wissenschaften und zugleich vieldeutig in der Literatur und Poesie. Wissenschaftliche Theorien arbeiten häufig mit Wörtern, Bildern und Metaphern, die der Alltagssprache entstammen. Eine lebendige Sprache ist in der Lage, sprachschöpferisch Neues zu bezeichnen. So kann sich Wissenschaft der Gesellschaft mitteilen und umgekehrt. Wird diese Verbindung gekappt, können die Wissensteilhabe und die gesellschaftliche Legitimation schnell schwinden. Es wäre sicher von Interesse, die Wechselbeziehungen zwischen Sprache und wissenschaftlicher Erkenntnis näher zu untersuchen. Je weniger in der Wissenschaft deutsch gesprochen wird, desto weniger wird die Gesellschaft über Wissenschaft sprechen.

Die Bedeutung der deutschen Sprache ist durch die politischen Ereignisse im 20. Jahrhundert, insbesondere nach dem Ersten Weltkrieg und durch die Gräueltaten während der nationalsozialistischen Zeit sehr zurückgegangen.

Aus Sprachliebe wurde Sprachscham und im Ausland war Deutsch die Sprache der Unmensen. Es war ein langer Weg, bis Deutsch in der Welt wieder gleichgesetzt wurde mit Kultur, Wissenschaft, Zivilisation und Völkerverständigung. Die Deutschen selbst haben nach Überwindung ihrer Sprachscham dann zu einer Gleichgültigkeit gegenüber ihrer Sprache gefunden. Aber es ist mit der Sprache ähnlich wie mit anderen Kulturgütern: Mangelnde Aufmerksamkeit macht sie weniger attraktiv, macht sie weniger reich und ausdrucksstark. Der Status der Sprache sinkt und es reduziert sich das, was man den Ausbau der Sprache nennt. Deshalb muss uns beschäftigen, was die Gleichgültigkeit mit uns macht! Sie beschädigt die Glaubwürdigkeit unserer Sprachpolitik. Die Sprachpolitik in Deutschland hat ihre direkte Auswirkung auf die Sprachvermittlung im Ausland. Die deutsche Sprache ist kein Selbstläufer, man muss in sie investieren und eine aktive Sprachpolitik betreiben. Und man muss bei den jungen Menschen anfangen.

Es geht nicht darum, zum x-ten Mal mit großer Leidenschaft allgemeine Erklärungen zu formulieren oder Appelle zu veröffentlichen, die Leidenschaft für die deutsche Sprache muss sich in praktischen Maßnahmen und konkreten Initiativen zeigen.

Als Goethe-Institut ist es uns ein besonderes Anliegen, das Zusammenspiel von Innen und Außen zu verhandeln. Wenn wir in Deutschland die Verantwortung für die eigene Sprache ernst nehmen, dann erleichtern wir auch den Deutschlehrern, den Sprachdozenten an Universitäten und vielen, die sich im Ausland aus Enthusiasmus für die Fremdsprache Deutsch einsetzen, das Geschäft der Sprachvermittlung. So kann glaubwürdig gearbeitet werden.

Es lässt sich inzwischen feststellen, dass eine neue Aufmerksamkeit und Zuwendung der deutschen politischen, kulturellen und gesellschaftlichen Öffentlichkeit gegenüber der deutschen Sprache eingesetzt hat. In Deutschland hat man aufgrund der verstärkten Zuwanderung begriffen, dass der Schlüssel zur Integration die deutsche Sprache ist. Die entsprechenden Voraussetzungen dafür zu schaffen ist eine entscheidende Bedingung für eine offene und verantwortliche Gesellschaft. Das hat auch für die Deutschen den Wert ihrer Sprache verdeutlicht.

Im Ausland erleben wir weltweit eine Zunahme der Deutsch Lernenden. 100 Millionen Menschen sprechen Deutsch als Muttersprache und noch einmal so viele als Fremdsprache. Aktuell lernen derzeit 15,4 Millionen Menschen Deutsch. Allein bei den rund 160 Goethe-Instituten in der Welt hatten wir in den letzten

fünf Jahren eine Steigerung von 20%, bei den Deutschprüfungen um 30%. Für neue Zugänge zur deutschen Sprache im Ausland setzen sich neben dem Goethe-Institut als Hauptvermittler auch die Zentralstelle für das Auslandsschulwesen und der DAAD ein. Das Interesse an der deutschen Sprache ist immer besonders groß, wenn ihr fachlicher und beruflicher Nutzen erkennbar ist oder wenn sich kulturelles Interesse auf bestimmte Entwicklungen fokussieren lässt. Deutschland als Wissenschafts- und Bildungsstandort hat eine hohe Attraktivität.

Ein zusätzlicher Faktor zur Zunahme von Deutschlernenden wird sicher durch das künftige Fachkräftezuwanderungsgesetz ausgelöst werden. Schon jetzt bietet das Goethe-Institut in Südostasien und Südosteuropa Deutschkurse für Pflegepersonal an.

Insgesamt lässt sich feststellen, die Tendenz ist in den meisten Ländern steigend, in einigen konstant. Es gibt auch ein paar Ausreißer nach unten, besonders deutlich war das bei Russland. Das war einem besonderen Umstand geschuldet. Bis vor einigen Jahren war Deutsch die erste Fremdsprache in den Schulen. Sie wurde durch Englisch per Dekret abgelöst. Das führte russlandweit zu einem dramatischen Einbruch. Ende 2015 sprach sich das russische Bildungsministerium dafür aus, dass jeder Schüler zukünftig zwei Fremdsprachen lernt. Das Goethe-Institut unterstützte dieses Ziel mit einer Kampagne „Deutsch: die erste Zweite“. Die eng miteinander verzahnten Module decken die gesamte Biografie von Deutschlernern im Alter von acht bis 17 Jahren ab und entsprechen den russischen Bildungsstandards. 2017 schlossen sich dieser Initiative 38 Regionen an und durch Kooperationsverträge mit den Bildungsverwaltungen und Lehrerfortbildungsinstituten verbindlich. Damit konnte der Trend umgekehrt werden. Wir schreiben wieder steigende Zahlen. Unterstützt wurden die Maßnahmen noch durch ein Jahr der deutschen Sprache. Solche Kampagnen werden auch in anderen Ländern durchgeführt, beispielsweise in den USA, ein Land, das wenig in Fremdsprachen investiert und in dem Deutsch als Fremdsprache keinen einfachen Stand hat. Die Germanistiklehrstühle an den Universitäten oder Lehrbeauftragte für Deutsch als Fremdsprache sind unter Druck geraten. Das Goethe-Institut hat im Oktober 2018 als Projektleiter gemeinsam mit dem Auswärtigen Amt und dem Bundesverband der Deutschen Industrie ein Deutschlandjahr gestartet, das mit 1000 Veranstaltungen in allen 50 US-Bundesstaaten präsent sein wird. Dabei spielt auch die deutsche Sprache eine wichtige Rolle. Als fahrbares Werbe-Vehikel für die deutsche Sprache tourt ein Wanderbus quer durch die USA. Er wird an 60 Highschools und Universitäten Station machen. An Bord sind Lern- und Bildungsprofis, die mit informativen

Präsentationen Schülerinnen und Schüler dazu motivieren, erste Erfahrungen mit der deutschen Sprache zu sammeln. Es gibt zusätzlich ein erfolgreiches Schüleraustauschprogramm GAPP (German American Partnership Program) bei dem jährlich etwa 6.000 Schülerinnen und Schüler in beide Richtungen gehen.

Interessant in der osteuropäischen Region ist ein Projekt, das trotz der politischen Spannungen ein großartiger Erfolg wurde. Es ist die Internationale Digitale Netzwerkuniversität, an der sich zehn Universitäten aus Georgien, Russland, der Ukraine, Österreich und Deutschland beteiligen. Das Auswärtige Amt unterstützt das Projekt im Rahmen des Programms „Östliche Partnerschaften“. Wissenschaft kann mit digitalen Anwendungen die physischen Ländergrenzen überwinden. Gegen den politischen Druck haben sich die Universitäten behauptet. Ein Zeichen für internationale Zusammenarbeit und Konfliktprävention und ein Loch in der Mauer. Begonnen hat dieses weltweit einmalige Projekt, das überwiegend in deutscher Sprache geschieht, im Wintersemester 2018/19. Die Einschreibungen sind beeindruckend. Es soll eine dauerhafte Zusammenarbeit sein, die möglichst noch erweitert werden soll. Die Prüfungen werden regulär angerechnet. Die Digitale Netzwerkuniversität will mit Weltoffenheit, Verantwortung und kultureller Vielfalt einen Gegenentwurf zu nationalistischen Tendenzen sowie wissenschaftliche Erkenntnisse gegen alternative Fakten setzen. Betreut wird das Projekt vom Goethe-Institut Moskau.

Ein weltweites Erfolgsmodell sind die sogenannten PASCH-Schulen, „Schulen: Partner der Zukunft“, vor zehn Jahren durch das Auswärtige Amt initiiert. Das Goethe-Institut unterstützt von den rund 2.000 PASCH-Schulen etwa ein Drittel, die anderen werden von der Zentralstelle für das Auslandsschulwesen betreut. Es werden in den Schulen der Gastländer deutsche Sprachabteilungen aufgebaut, Lehrer fortgebildet und die Schulen mit Lehr- und Lernmaterial ausgestattet. Es gibt für die besten Schüler Sommercamps in Deutschland und über eine Online-Community bleiben sie in Kontakt. Es ist mehr als nur Deutschlernen, es ist begeistern und Perspektiven schaffen. Für die Lehrkräfte und für die Schulleiter/innen gibt es Fortbildung und PASCH-Konferenzen. Schwerpunkte sind unter anderem China, Indien, Indonesien und Brasilien. Damit werden richtig große Zahlen gemacht! Wie hoch die Lernmotivation ist, kann man an der Internationalen Deutscholympiade erleben, weltweit der größte Wettbewerb der deutschen Sprache. Sie findet alle zwei Jahre statt, zuletzt im Juli 2018 in Freiburg, organisiert vom Goethe-Institut und dem Internationalen Deutschlehrerverband. Ausgezeichnet werden Lehrer für besonders innovativen Unterricht und Schüler für ihre überzeugenden Präsentationen in drei Sprachniveaus.

Deutsch lernen kann man auch in Deutschland. Hier sind die 12 Goethe-Institute Marktführer für Sprachreisen für Erwachsene und für Prüfungen. Der Erfolg muss jedoch immer wieder neu errungen werden. Es gibt gerade in dem Anwendungsgebiet der Sprachvermittlung erhebliche Veränderungen. Dazu zählen der gestiegene Bedarf an Lehrkräften, das Lehren und Lernen mit digitalen Medien und nicht zuletzt die gestiegenen Qualitätsanforderungen.

Das Goethe-Institut verfügt mit dem Grünen Diplom und dem Grünen Fortbildungsportfolio über ein eigenes Aus- und Fortbildungsprogramm, das weltweit einheitlich und standardisiert für Lehrkräfte im Ausland durchgeführt wird. Auf diese Weise sind unsere Lehrkräfte in der Lage, Deutsch auf medial und methodisch modernste Weise zu unterrichten. Digitale Medien werden konsequent eingesetzt. Zwischen 2006 und 2018 haben insgesamt ein Drittel der Lehrkräfte das Grüne Diplom an Goethe-Instituten im Ausland absolviert. 2018 kamen 1.600 Lehrerinnen und Lehrer aus der ganzen Welt nach Deutschland, um ihre Lehrkompetenz zu erweitern. Die Programme zur Lehrerfortbildung müssen deutlich ausgeweitet werden. Dabei spielt das Fortbildungsprogramm „Deutsch Lehren Lernen“ eine große Rolle. Hier sind insbesondere die Online-Fortbildungsprogramme zu betonen.

Digitale Plattformen sind für das Goethe-Institut immer stärker im Fokus. In über 10.000 Kursen auf der Lernplattform sind aktuell 120.000 Deutschlerner in Sprachkursen aktiv. Aber auch Spiele und Games zählen zu Lernpfaden, die wir mit Erfolg und mit Freude am Lernen einsetzen. Die erste Lernspielapp „Das Geheimnis der Himmelscheibe“, die 2011 erschien, wurde mehr als eine Million Mal heruntergeladen. Lernerinnen und Lerner werden zu Netzagenten und lernen spielerisch die Lebenswirklichkeit von deutschen Jugendlichen kennen – und lernen Deutsch. Es gibt auch berufsorientierte Bausteine für Mediziner, Altenpflege usw. Diese Bausteine werden durch das künftige Fachkräftezuwanderungsgesetz schrittweise erweitert. Als Erstes soll 2019 ein innovatives Anwerbe- und Fortbildungsprogramm dazu beitragen, weltweit neue Lehrkräfte zu gewinnen. Schnelle Qualifizierung ohne Qualitätsverlust bei der Ausbildung – das ist hier die Herausforderung. Ein zweiter Schwerpunkt betrifft die Intensivierung von Programmen für Hochqualifizierte wie zum Beispiel die „Studienbrücke“, die gemeinsam mit dem DAAD Studieninteressierte in der Schule im Herkunftsland auf ihr Studium in Deutschland vorbereitet und den erfolgreichen Absolventinnen und Absolventen Kontingentplätze an Partneruniversitäten in Deutschland garantiert. Zusätzlich wird zur Zeit am Aufbau eines digitalen Studienkollegs gearbeitet,

das nach Schulabschluss im Herkunftsland auf ein Studium in Deutschland vorbereitet. Um ein breit aufgestelltes, weltweit abrufbares Potenzial deutschsprechender Fachkräfte mittel- und langfristig aufzubauen, ist auch eine Verstärkung der Förderung von Deutschunterricht an öffentlichen Schulen erforderlich, hier insbesondere an Berufsschulen.

Die Qualität der Sprachvermittlung ist einerseits abhängig von methodischen, didaktischen und technischen Bedingungen, andererseits ist der Ausdrucksreichtum einer Sprache durch ihren Wortschatz gegeben, der im Deutschen so umfassend und differenziert ist wie in kaum einer anderen Sprache. So hat eine hier an der BBAW vor einigen Jahren entstandene Untersuchung gezeigt, dass in dem zu Grunde gelegten umfangreichen Corpus der deutschen Gegenwartssprache mehr als fünf Millionen verschiedener Wörter vorkommen. Das ist fast ein Drittel mehr, als in einem vergleichbaren Corpus von vor hundert Jahren belegt sind. Das zeigt, wie dynamisch sich der deutsche Wortschatz in den letzten hundert Jahren entwickelt hat und sicherlich weiter entwickeln wird. Da die deutsche Sprache – wie wir gehört haben – über keine Zentralinstanz verfügt, die über die Reinheit der Sprache wacht, ist es umso wichtiger, verlässliche Informationen über Schreibweise, Aussprache, Bedeutungen, grammatische Eigenschaften, regionale wie stilistische Variationen zu haben. Sie gibt den Lehrenden und den Lernenden erst die Sicherheit über eine korrekte Qualität. Haben wir doch, werden vielleicht einige von Ihnen wissen: der *Duden*, der *Wahrig*, das *Grimm'sche Wörterbuch*. Das war einmal. Die Verlage können diese großartige Tradition aus wirtschaftlichen Gründen nicht mehr fortsetzen, die Lexikabteilungen sind größtenteils geschlossen. Und das *Grimm'sche Wörterbuch* mit seinen 32 Foliobänden und den rund 350.000 Wörtern, deren Bedeutung es beschreibt und mit Belegen versieht, war bereits überholt, bevor die jeweiligen Bände erschienen sind. Ohne eine angemessene Darstellung des Wortschatzes wird die Sprache verwildern und keine standardisierten Bezugspunkte mehr bieten.

Das zu leisten, ist nicht mehr in Form eines klassischen gedruckten Wörterbuchs möglich, und sei es auch noch so umfangreich. Der Grimm hat zu seiner Bearbeitung über hundert Jahre erfordert. Das Ergebnis ist ein wissenschaftliches Meisterwerk. Aber so beeindruckend der Grimm auch ist – er umfasst nicht einmal ein Zehntel der Wörter, die im Gegenwartsdeutschen vorkommen. Und er beschränkt sich auf knappe Bedeutungsangaben und zahlreiche Belege; viele andere Eigenschaften der Wörter, beispielsweise die Aussprache, die Flexionsweise oder die grammatischen Merkmale, werden gar nicht behandelt. Sie zu kennen ist aber beispielsweise

für einen Deutschlerner nicht weniger wichtig als die Bedeutung. Hier muss man methodisch anders vorgehen, nicht dadurch, dass man die Sorgfalt und die Expertise des Lexikographen preisgibt, sondern indem man sie durch die Mittel ergänzt, die uns die digitalen Techniken an die Hand geben. So wird es auch möglich, rasch auf Veränderungen der Sprache zu reagieren, und ebenso wird es möglich, sich dem Bedarf und den Wünschen der unterschiedlichen Nutzer anzupassen. Ziel muss es daher sein, ein digitales Informationssystem zu entwickeln und zu betreiben, das den deutschen Wortschatz umfassend und verlässlich beschreibt und nutzerfreundlich anbietet, dabei auf die laufende Aktualisierung der Sprachveränderungen achtet und auch die bereits vorhandenen Wörterbücher einarbeitet. Dieses digital verfügbare lexikalische System muss über das Internet jedermann zugänglich sein. Sinnvoll ist es, durch modularen Aufbau verschiedene Forschungsstätten zu beteiligen und auch Nutzer selbst in die Entwicklung einzubinden. Die Berlin-Brandenburgische Akademie der Wissenschaften hat bereits mit dem Digitalen Wörterbuch der Deutschen Sprache entscheidende Vorarbeiten geleistet. Es muss jedoch erheblich ausgebaut werden. Dazu eignet sich für die Aufbauphase eines Zentrums für digitale Lexikographie der deutschen Sprache die Schirmherrschaft der Union der deutschen Akademien. In der Aufbauphase sollten bereits feste Kooperationen mit einer Reihe von Einrichtungen eingegangen werden, die auf dem Gebiet tätig sind, sei es, weil sie selbst lexikographisch arbeiten oder weil sie für die Nutzung der lexikalischen Ressourcen von Bedeutung sind. Das könnten sein: Das Leibniz-Institut für Deutsche Sprache in Mannheim, die Deutsche Akademie für Sprache und Dichtung in Darmstadt und das Goethe-Institut in München. Aus meinen Ausführungen zur Sprachvermittlung im Ausland ist sicher deutlich geworden, wie entscheidend die Vermittlungsqualität von einem gesicherten Wortschatz abhängt. Wir haben diese Entwicklung eines solchen Zentrums in den Instrumentarien des Goethe-Instituts bereits berücksichtigt, wir haben auch bereits bei einigen Programmen Schnittstellen zum Zentrum für digitale Lexikographie vorgesehen, etwa bei der digitalen Lernplattform oder bei der Digitalen Netzwerkuniversität. Damit können bessere Wortschatzhilfen für die Onlinekurse bereitgestellt werden. Für das Zentrum seinerseits eröffnet sich mit dieser Kooperation der Zugang zu einer der wichtigsten Nutzergruppen von Wörterbüchern – den Deutschlernern aus aller Welt. Lassen Sie mich die Vorteile dieser Kooperation für den Deutschunterricht im In- und Ausland, aber auch für die Wörterbuchmacher selbst an drei kurzen Beispielen erläutern:

1. In dem digitalen Wörterbuch sind die Informationen zu einem Wort und die Belege, auf denen diese Informationen beruhen, direkt miteinander

verbunden; jeder kann sich über das Internet beides nach Belieben ansehen. Auch wenn daher ein Wort noch nicht umfassend, ja, vielleicht sogar überhaupt noch nicht beschrieben ist, kann sich jeder ein gutes Bild von seiner Verwendung und damit seiner Bedeutung, seiner üblichen Schreibweise und seinen grammatischen Eigenschaften machen. Das ist für Deutschlerner ebenso wichtig wie beispielsweise für Übersetzer.

2. Im digitalen Wörterbuch wird die Aussprache nicht durch eine Lautschrift umschrieben, die in der Praxis immer nur eine Annäherung bieten kann, sondern durch professionelle Sprecher eingesprochen. Das gibt den Deutschlernenden in aller Welt nicht nur eine weitaus bessere Vorstellung, wie ein Wort tatsächlich klingt, sondern macht es auch möglich, es überprüfbar nachzusprechen und die eigene Aussprache mit dem Original zu vergleichen.
3. Nutzer in aller Welt können jederzeit Lücken, Fehler und Unstimmigkeiten melden sowie eigene Korrekturvorschläge machen und so wesentlich zur Verbesserung des Wörterbuchsystems beitragen. Anders als beim klassischen Wörterbuch können diese Anregungen nämlich sofort zur Korrektur und zur Ergänzung berücksichtigt werden.

Das Zentrum für digitale Lexikographie der deutschen Sprache kann so die große deutsche Wörterbuchtradition mit digitalen Möglichkeiten fortsetzen. Es ist eine große Chance für die deutsche Sprache. Dieses Werkzeug wird noch wichtiger werden als Voraussetzung für automatische Sprachverarbeitung, auf die wir uns im Zuge der Digitalisierung vorbereiten müssen. Nicht zuletzt kann es, da in seinem Umfang nicht beschränkt, schrittweise immer weiter ausgebaut werden und so auf lange Sicht die deutsche Sprache, die ja nicht nur in der Bundesrepublik Deutschland gesprochen wird, allen Interessierten in seiner ganzen Vielfalt erschließen.

Johann Wolfgang von Goethe soll bei diesem Thema das letzte Wort haben. Er schreibt in seinen *Maximen und Reflexionen*: „Die Muttersprache zugleich reinigen und bereichern ist das Geschäft der besten Köpfe. Reinigung ohne Bereicherung erweist sich oft als geistlos; denn es ist nichts bequemer als von dem Inhalt abzu- sehen und auf den Ausdruck zu passen. Der geistreiche Mensch knetet seinen Wortstoff, ohne sich zu bekümmern, aus was für Elementen er bestehe; der geistlose hat gut rein sprechen, da er nichts zu sagen hat.“





berlin-brandenburgische  
AKADEMIE DER WISSENSCHAFTEN



Akademie der Wissenschaften  
zu Göttingen



Sächsische Akademie der Wissenschaften  
zu Leipzig



Akademie der Wissenschaften  
und der Literatur | Mainz

IDS

LEIBNIZ-INSTITUT FÜR  
DEUTSCHE SPRACHE



UNION  
DER DEUTSCHEN AKADEMIEN  
DER WISSENSCHAFTEN

Das Zentrum für digitale Lexikographie der deutschen Sprache (ZDL) hat am 1. Januar 2019 seine Arbeit aufgenommen. Am 29. Januar 2019 fand im Leibniz-Saal der Berlin-Brandenburgischen Akademie der Wissenschaften der feierliche Auftakt statt, zu dem rund 250 Gäste zusammengekommen sind. Um diesem Tag ein kleines Denkmal zu setzen, haben wir diesen Band zusammengestellt. Er enthält alle Vorträge, die während des Abends präsentiert worden sind.

ISBN: 978-3-939818-87-8

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung