



Akademie der Wissenschaften
zu Göttingen

„Gewöhnlich glaubt der Mensch, wenn er nur Worte hört, es müsse sich dabei doch auch was denken lassen.“

Johann Wolfgang von Goethe (1749 - 1832)

Sonderausgabe

Akademie digital

Geistes- und
Naturwissenschaften

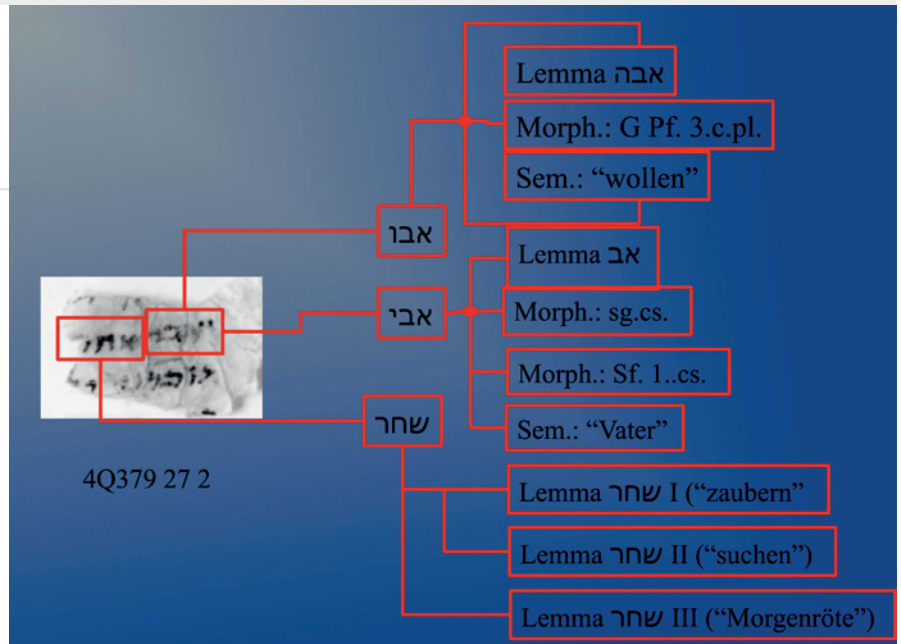
Kompetenz durch
Kooperation

Sehr geehrte Damen und Herren,

die Entwicklung der digitalen Möglichkeiten eröffnet nicht nur der Wissenschaft, sondern auch einem breiteren Kreis von Interessenten völlig neue und bisher ungeahnte Zugänge zu Textkorpora, auf die bis vor kurzem selbst Fachwissenschaftler zuweilen nur eingeschränkten oder mühsamen Zugriff hatten. Digitale Korpora wie z.B. die von Gregory R. Crane initiierte berühmte Perseus Digital Library (<http://www.perseus.tufts.edu/hopper/>) erlauben nicht nur einen bequemen Zugang zu einer Großzahl von Texten, sondern auch detaillierte Recherchen, für die sich frühere Generationen mühsam eigene Erschließungswerkzeuge wie den legendären „Zettelkasten“ selbst erstellen mussten. Wird damit aber nicht auch das Wörterbuch als ein klassisches Ergebnis dieser Arbeiten am Zettelkasten schlicht obsolet?

Dieser Frage gingen **Dr. Volker Harm** und **PD Dr. Ingo Kottsieper**, Mitarbeiter zweier Wörterbuchprojekte der Akademie der Wissenschaften zu Göttingen, im Rahmen eines Workshops nach, den die Arbeitsgemeinschaft „Elektronisches Publizieren“ der Union der Akademien im Oktober 2013 in Berlin veranstaltete.

Ihr gemeinsamer Beitrag wird im Folgenden in gekürzter Form abgedruckt.



Ein einziges Wort aus einem Textfragment erlaubt verschiedene sprachliche Deutungen. Die Datenbank des digitalen Qumran-Wörterbuchs ist entsprechend komplex.

Alter Wein in neuen Schläuchen?

Wörterbuchprojekte im Zeitalter digitaler Korpora

Von **Dr. Volker Harm** und
PD Dr. Ingo Kottsieper

Gegenwärtig vollziehen sich sowohl im Bereich der elektronischen Korpora als auch in der Lexikographie beachtliche Umbrüche: Die Zahl, der Umfang und auch die Qualität der digitalen Textsammlungen steigt stetig an, und bei den Wörterbüchern geht der Bedeutungsverlust angestammter Autoritäten wie des Duden oder Brockhaus mit dem Aufblühen einer vielfältigen Landschaft von Internetwörterbüchern und Crowd-Sourcing-Projekten einher. In dieser Lage ergibt sich auch für die Wörterbuchvorhaben der Akademien die Notwendigkeit einer Positionsbestimmung. Denn je besser Korpora

durchsucht und ausgewertet werden können, desto geringer erscheint die Bedeutung eines Wörterbuchs, das sich auf das Korpus bezieht: Wozu bedarf es noch der Vermittlungsleistung eines Lexikographen, wenn ein Nutzer alle ihn interessierenden Fragen unmittelbar am Korpus beantworten kann? Es stellt sich also – überspitzt formuliert – die Frage, wozu künftig noch Wörterbücher (sei es gedruckte oder digitale) benötigt werden, wenn die Informationen, die sie enthalten, auch direkt aus den digitalen Korpora gewonnen werden können. Oder reichen hier möglicherweise sogar Wörterbücher aus, die nicht von Fachleuten erarbeitet, sondern durch die „Schwarmintelligenz“ von Semi-Spezialisten oder inter-



Dr. Volker Harm



PD Dr. Ingo Kottsieper

essierten Laien zusammengetragen werden? Die Positionsbestimmung, zu der diese Fragen auch die Akademienlexikographie herausfordern, soll hier aus dem Blickwinkel zweier lexikographischer Vorhaben versucht werden, die sowohl hinsichtlich ihres Gegenstandes als auch ihrer digitalen Entwicklungsstufe unterschiedlicher kaum sein können: aus der Sicht des traditionsreichen Deutschen Wörterbuchs von Jacob Grimm und Wilhelm Grimm, dessen Korpusbasis ein umfangreiches Zettelarchiv bildet, und des 2002 begründeten, vollständig digital erstellten Qumran-Wörterbuchs. Zunächst soll hier auf das Verhältnis von Wörterbuch und Wörterbuchkorpus in der Neubearbeitung des Grimm eingegangen werden (²DWB). Das Vorhaben soll die am stärksten veralteten Teile der von 1854 bis 1960 in 32 Bänden erschienenen Erstauflage des Wörterbuchs ersetzen, d. h. die Buchstabenstrecke A-F, die noch von den Brüdern Grimm selbst erarbeitet worden ist. Wie die Erstbearbeitung, beschreibt auch die Neubearbeitung den in der neuhochdeutschen Schriftsprache von der Mitte des 15. Jhs. bis zur Gegenwart gebräuchlichen Wortschatz, einschließlich seiner Vorgeschichte im Alt- und Mittelhochdeutschen. Im Mittelpunkt der Darstellung steht dabei die Bedeutungsgeschichte der Wörter. Einer der wesentlichen konzeptionellen Grundpfeiler der Neubearbeitung ist die konsequente Korpusbasierung: Alle lexikogra-

phischen Befunde – zur Bedeutung, Grammatik, Schreibung sowohl des vergangenen als auch des gegenwärtigen Deutschen – sind aus dem Belegarchiv zu erarbeiten. Sekundäre Informationen wie etwa Befunde aus der Forschungsliteratur oder aus anderen Wörterbüchern spielen demgegenüber eine deutlich untergeordnete Rolle. Der konsequenten durchgängigen Korpusbasierung entsprechend ist auch das Belegarchiv des ²DWB verhältnismäßig umfassend angelegt: Es beruht auf einer systematischen Exzerption von ca. 8.000 Quellen, die vom 8. bis ins 20. Jh. reichen und enthält insgesamt rund 5,5 Mio. Belegzettel für den Buchstabenbereich A-F. Das ²DWB basiert damit auf einer wesentlich umfangreicheren und besseren Materialgrundlage als der alte Grimm, der nur rund 2,5 Mio Belegzitate für A bis Z umfasste.

Das ²DWB versteht sich nun nicht nur als historisches Bedeutungswörterbuch, sondern auch als Belegwörterbuch. Was damit gemeint ist, wird unmittelbar anschaulich, wenn man eine beliebige Seite des Wörterbuchs aufschlägt: Die Präsentation von Textausschnitten nimmt den mit Abstand größten Druckraum ein, Bedeutungserläuterungen und sonstige Kommentare beanspruchen demgegenüber erheblich weniger Platz. Der hohe Anteil der Belegzitate am Wörterbuchtext ist indes nicht nur ein Erbe der Brüder Grimm, die in den ausgewählten Textstellen gewissermaßen die Sprache

selbst zu Wort kommen lassen wollten, sondern erfüllt gleich eine ganze Reihe von Aufgaben: Neben der grundlegenden Funktion der Beglaubigung der semantischen und grammatischen Angaben im Wörterbuchartikel soll die abgedruckte Belegauswahl auch eine Art Fenster in das Wörterbucharchiv sein, dessen Strukturen und Verteilungen auf diese Weise im Ansatz sichtbar gemacht werden können. Damit ist diese Miniaturversion des Archivbefundes auch ein Angebot an den Nutzer, die Belege als Basis für eigene Recherchen zu verwenden, auch für solche, die nicht unmittelbar mit der Wörterbuchdarstellung zu tun haben. Das ²DWB bildet damit in gewisser Weise bereits für sich genommen eine Art Verbund von Korpus und Wörterbuch, wie er im digitalen Zeitalter als Idealfall angesehen wird.

Wenn der Belegteil des gedruckten Wörterbuchartikels ein Fenster in das Belegarchiv darstellt, so gerät, um im Bild zu bleiben, damit nur ein sehr kleiner Teil des Archivs überhaupt in den Blick des Nutzers. Das Verhältnis zwischen dem im Archiv vorhandenen Material und den im Wörterbuch ausgesuchten Belegen und Stichwörtern ist durch ein deutliches Ungleichgewicht gekennzeichnet: So entspricht das im Wörterbuch abgebildete Belegmaterial nur 5-6% des gesamten Archivmaterials, und von den im Archiv bezeugten Wortansätzen (es sind rund 300.000) werden durchschnittlich nur rund 12% in einem eigenen Wort-

artikel beschrieben. Ein sehr großer Teil der Textexzerpte und Wörter bleibt somit für den Benutzer unsichtbar. Für die anstehende Retrodigitalisierung der bisher nur in Druckform vorliegenden Bände des ²DWB bedeutet dies, dass neben das digitale Wörterbuch eigentlich auch eine Digitalisierung des Zettelarchivs zu treten hätte. Erst diese würde die Forschungsarbeit, die in dem Projekt geleistet wird, in ihrer ganzen Breite zugänglich machen. Als erster Schritt in Richtung auf eine solche umfassende Digitalisierung kann immerhin die vollständige Liste der Archivwörter angesehen werden, die auf der Homepage der Göttinger DWB-Arbeitsstelle einsehbar ist.

Der große Umfang des ²DWB-Archivs macht auch die verhältnismäßig langen Bearbeitungszeiten des Vorhabens verständlich: Die Erstauflage des DWB hatte rund 120 Jahre gebraucht, für die Neubearbeitung A-F werden zum Laufzeitende 2016 nochmals gut 55 Jahre Bearbeitungszeit zu Buche stehen. Aus heutiger Sicht mag hier der Gedanke naheliegend erscheinen, dass eine digitale Bearbeitung – vielleicht sogar durch eine „Crowd“ – viel Zeit und Geld gespart hätte und für künftige Projekte in der Tradition des Grimmschen Wörterbuchs auch sparen wird. Dieser Gedanke, so ‚angesagt‘ er angesichts erfolgreicher Online-Enzyklopädien ist, verfehlt indes den kritischen Punkt: Das Erstellen eines Wörterbuchs, das Bedeutungsentwicklungen von Wörtern aus 1200 Jahren deutscher Sprachgeschichte zum Gegenstand hat, ist und bleibt eine Arbeit, die im Kern nur von Philologen geleistet werden kann. Im Unterschied zur Arbeit an einer Enzyklopädie handelt es sich bei Sprachwörterbüchern auch nicht um eine Sammlung von mehr oder weniger guten Referaten über einen gegebenen Forschungsstand, sondern um Grundlagenforschung. Diese kann zwar durch digitale Werkzeuge unterstützt und bis zu einem gewissen Maß auch beschleunigt werden, sie kann durch diese Instrumente aber nicht ersetzt werden, und sie ist auch nicht – oder zumindest nicht in einem vernünftigen Verhältnis von Kosten und Nutzen – durch die Kollaboration einer großen Zahl von Semi-Spezialisten zu leisten.

Dies wird auch im Hinblick auf das Qumran-Wörterbuchprojekt der Göt-



Ein Zettelkasten in der Göttinger Arbeitsstelle des Deutschen Wörterbuchs

tinger Akademie deutlich, das unter der Leitung von Prof. Dr. Reinhard Kratz vor elf Jahren seine Arbeit aufgenommen hat und von seinem Charakter her dem ²DWB völlig entgegengesetzt ist. Obwohl es als von Beginn an digital angelegtes Projekt die Vorteile illustriert, die ein digital erschlossenes Korpus bietet, wird auch hier deutlich, dass solche Korpora das klassische Wörterbuch nicht einfach ersetzen können.

Im Rahmen dieses Projektes soll ein umfassendes Wörterbuch aller hebräischen und aramäischen Texte der hellenistisch-römischen Zeit erstellt werden, die in ihrer Mehrzahl in Qumran gefunden wurden. Das Korpus dieser Texte ist im Gegensatz zu dem des ²DWB überschaubar und kann zu 100 Prozent einbezogen werden – die fragmentarischen Texte bieten nur etwa 90.000 Wörter, die soweit erhalten sind, dass sie für das Wörterbuch sinnvoll herangezogen werden können. Der bekannte fragmentarische Zustand der Quellen begründet weitere Unterschiede zum Korpus des ²DWB: Auf Grund des häufig nicht erhaltenen Kontextes und durch das vokallose und damit potentiell mehrdeutige Schriftsystem können einzelne Wörter sprachlich mehrdeutig sein. Darüber hinaus finden sich in der Sekundärliteratur unterschiedliche Lesungen und Interpretationen der Fragmente, die bei der Wörterbucharbeit berücksichtigt werden müssen. Das dadurch bedingte komplexe Geflecht von unterschiedlichen Lesungen und Deutungen wird im Rahmen

dieses Projektes durch eine umfassende Datenbank abgebildet. Diese enthält nicht nur alle Lesevorschläge und die daraus resultierenden sinnvollen Interpretationen hinsichtlich der grammatikalischen Form und Bedeutung, sondern auch die syntaktische Einordnung der einzelnen Wörter in Phrasen und Sätzen. Darüber hinaus findet der Nutzer hier sämtliche Informationen zu Parallelversionen nicht nur in den unterschiedlichen Manuskripten, sondern auch in Zitaten innerhalb anderer Texte. Mithin wird nach Abschluss der Arbeiten ein vollständig digital erschlossenes Textkorpus vorliegen, das verspricht, alle Fragen hinsichtlich der Sprache der Texte und ihrer Textgrundlage beantworten zu können.

Dennoch macht selbst eine solche umfassende digitale Repräsentation des Korpus weder die lexikographische Arbeit noch das Wörterbuch überflüssig. Grundsätzlich gilt, dass ein nicht semantisch erschlossenes Korpus nur demjenigen zugänglich ist, der die von diesem Korpus repräsentierten Sprachstufen beherrscht. Nur dieser ist in der Lage, die möglichen Bedeutungsnuancen eines Lemmas aus dem gegebenen Kontext zu erschließen. Es hilft dem Nutzer wenig, sämtliche sicheren oder möglichen Belege eines Lemmas präsentiert zu bekommen, wenn er diese nicht oder nur teilweise versteht. Dies bedeutet umgekehrt, dass nur ein semantisch erschlossenes Korpus über die engen Grenzen des Kreises der Spezialisten von Nutzen ist.



Unschärfe Suche

Weniger Suchoptionen
Bibliothek: VZG | Anmelden | Abmelden | Datenbankmenü

Über die Datenbank
Willkommen

GBV - Literatur zur Wortforschung







Die Online-Datenbank *Literatur zur Wortforschung* ist im Rahmen des Akademievorhabens *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm (Neubearbeitung)* entstanden. Sie ermöglicht es Wissenschaftlern und interessierten Laien, durch Online-Recherche auf Forschungsliteratur zu einzelnen Wörtern zuzugreifen.

Startseite der Online-Datenbank „Literatur zur Wortforschung“. Die Datenbank ist ursprünglich als internes Hilfsmittel im Projekt „Deutsches Wörterbuch von Jacob und Wilhelm Grimm“ (Neubearbeitung) entstanden.

Dieser Sachverhalt wird besonders an solch speziellen Korpora deutlich wie dem, das dem Qumran-Wörterbuch zugrunde liegt. Ohne eine semantische Erarbeitung der einzelnen Belege wäre es nur für den ausgewiesenen Qumranspezialisten hilfreich, der die Hintergründe und Kontexte der Quellschriften im Detail kennt und daher auch Einzelbelege einordnen kann. Anderen Wissenschaftlern z.B. aus der Theologie oder der Semiotik – von interessierten Laien gar nicht zu sprechen – würde es dagegen kaum helfen, die zugrundeliegenden Quellen zu erschließen. Mit einer solchen semantischen Erschließung des Korpus ist jedoch schon ein wesentlicher Teil der klassischen Wörterbucharbeit geleistet.

Die Ergebnisse einer solchen semantischen Erschließung werden für den Nutzer aber nur dann verwendbar, wenn er diese im Überblick kennt bzw. aus ihnen auswählen kann. Um alle Belege des Lemma X mit der Bedeutung Y zu suchen, muss der Benutzer erst einmal wissen, welche Bedeutungen dem Lemma X zugeordnet werden können und dass Y dazu gehört. Folglich muss auf irgendeine Weise eine Übersicht über die Bedeutungskategorien geboten werden, was de facto nichts anderes ist als das Grundgerüst eines Wörterbuchartikels.

Vergleichbares gilt für Phrasen, deren Verständnis für die semantische Erschließung eines Lemmas und für das Verständnis von Texten grund-

legend ist: Will man dem Nutzer nicht zumuten, durch „trial and error“ selbst die Phraseologie eines Lemmas zu ermitteln, muss auch hier eine Auflistung der belegten Phrasen bzw. semantischen Kontexte eines Lemmas angeboten werden. Insbesondere bei Lemmata mit einer komplexen Bedeutungsstruktur und Phraseologie kommt es dem Interesse des Nutzers entgegen, wenn dies auch in einer vorsortierten und gewichteten Form angeboten wird.

Mithin setzt die semantische Erschließung auch eines digital zugänglichen Korpus die lexikalische Arbeit voraus. Darüber hinaus kann eine der wesentlichen Leistungen des klassischen Wörterbuches, nämlich eine Übersicht über die Bedeutungskategorien und die Phraseologie eines Lemmas zu bieten, auch einen praktikablen und komfortablen Zugang zu den Daten des Korpus ermöglichen. Der Verzicht auf das Wörterbuch würde für den Nutzer eines digital erschlossenen Korpus bedeuten, dass er sich die Antworten auf die Fragen, die ihm eigentlich das Wörterbuch beantworten könnte, selbst aus dem im Korpus gebotenen Daten erarbeiten muss.

Dieser letzte Aspekt illustriert aber zugleich den Mehrwert, den das digital erschlossene Korpus gegenüber dem klassischen Wörterbuch aufweist. Es bietet dem Nutzer die Möglichkeit, die den Entscheidungen des Wörterbuches zugrundeliegenden Daten in ihrer Gesamtheit zu sehen, und ermöglicht

es ihm auf diese Weise, diese Entscheidungen nachzuvollziehen und kritisch zu würdigen.

Somit ergibt sich, dass die Kombination von Wörterbuch und Korpus eine ideale Situation darstellt. Im Qumran-Wörterbuchprojekt wird dem dadurch Rechnung getragen, dass das Wörterbuch von vornherein in digitaler Form erstellt wird und dabei unmittelbar auf die Daten des Korpus zurückgreift. Die Angaben in den Wörterbuchartikeln sind mit den zugrundeliegenden Daten fest verlinkt. Einerseits können auf diese Weise die Artikel des Wörterbuches als komfortabler Zugang zu den weiterführenden Daten des Korpus genutzt werden. Andererseits kann man auch umgekehrt direkt aus dem Korpus auf die relevanten Artikel zugreifen, und sich so der Ergebnisse der lexikographischen Arbeit bedienen, die mehr ist als eine bloße Zusammenstellung von Belegstellen.

Wörterbuch und digitales Korpus ergänzen sich somit in hervorragender Weise und bilden zusammen eine wissenschaftlich höchst wertvolle Kombination. Durch die heute verfügbaren digitalen Techniken lassen sich beide auch zu einer Einheit verschmelzen, die die zugrundeliegenden Daten sowohl vollständig als auch in wissenschaftlich verantworteter Aufarbeitung dem Nutzer zur Verfügung stellen. Damit ergeben sich neue Perspektiven, die das klassische Wörterbuch alles andere als alt aussehen lassen.

Impressum: Dr. Angelika Schade, Generalsekretärin der Akademie der Wissenschaften zu Göttingen

Gestaltung: Adrienne Lochte, alochte1@gwdg.de, Akademie der Wissenschaften zu Göttingen, Theaterstraße 7, 37073 Göttingen